THE UNIVERSITY
of EDINBURGH

Chromatin Organisation in Breast Cancer

Sehrish Rafique[1,2]

For the Degree of Doctor of Philosophy

1. MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, Western General Hospital, Edinburgh, EH4 2XU, U.K.

2. Edinburgh Breakthrough Research Unit, University of Edinburgh, Crewe Road South, Edinburgh, Scotland, EH4 2XU, UK.

Declaration

I hereby declare that except where specific reference is made to other sources, the work contained in this thesis is the original work of the author. It has been composed by myself and has not been submitted, in whole or in part, for any other degree, diploma, or other qualification.

Sehrish Rafique April, 2013

Acknowledgements

I would first like to thank the Medical Research Council and Breakthrough Breast Cancer for funding my research during my PhD. I would like to thank my supervisor, Wendy Bickmore, for helping me see what it really is to be a good scientist and also for the opportunity to be a part of her lab. I would like to thank Nicholas Gilbert for always having the time for helpful discussion and for his contagious enthusiasm for chromatin and my project. Thanks are also due to my thesis committee members, Richard Meehan, Andy Sims and Dana Faratian, for their time, suggestions and encouragement during my progress meetings.

I would like to thank all members of the Bickmore lab past and present for helpful discussion, support and making the lab such an enjoyable environment to work in. In particular I would like to thank the following people for help with techniques and/or sharing reagents: Shelagh Boyle, Elizabeth Kerr, Pierre Therizols, Emmanuelle Deniaud, Ragnhild Eskeland and Ro Billingworth. Thanks also to everyone in the CGE section for keeping our weekly meetings so lively that I had nothing to fear during my viva and also for providing cake at every possible opportunity.

I would like to thank Duncan Sproul for his time and effort spent helping me get to grips with R and statistics at the start of my project. For this I would also like to offer thanks to Jennifer Huffman, Alasdair Jubb, James Reddington and Colin Semple. Many thanks also to Jeremy Thomas for providing me with breast tissue samples and to Lynne Johnstone for sectioning them for me.

Having spent many hours in the dark imaging slides or processing data I must give a special thanks to Paul Perry and Matthew Pearson who were always at hand for help with equipment/analysis and emergency (or not) cakes/biscuits/coffee/tea. Many thanks also to all the people in Core Services especially Pauline, Eilidh, Craig, Gary, Derek and Graeme.

On a more personal note I would like to acknowledge and thank: the Chromatin Crew (*Chief* Iain, Gillian, Sam and Adam) for enjoyable discussion of ideas; the Inner Circle (Kerrie-cool, Katykins, Juanito, Amy-bear, *SA*dam, Smills, Irish-G and Heidilicious) plus

the Outer Hemisphere (David), the MRC Sharks (black-shark, echo, the-fish, the-viking, the-superman, spanish-omlette), Victor, Cici and Raffaelle for their support and for showing me such an enjoyable time in Edinburgh! I would like to acknowledge the Mosque Kitchen and Redbox Noodlebar for keeping me from starving whilst I wrote my thesis.

Finally, I must also thank friends from home (particularly Safiyyah & Hollie) and my sisters (Henna, Faizah & Azimah) for their support and for keeping me grounded. The biggest thanks of all goes to my parents: Mum – thank you for the many years of prayers for one exam after another and particularly this final one; Dad – thank you for everything you went through with moving and making a life here, without which, none of this would have been possible.

Table of Contents

List of Figures

Figure 6.5: Transcription Correlation Scores (TCS) at the locus containing the protocadherin gene cluster on chr5q31.3

List of Tables

Abbreviations

2D	Two Dimensional
3C	Chromosome Conformation Capture
4C	Circular 3C
5C	3C Carbon Copy
5hmC	5- hydroxyl methyl cytosine
5mC	5- methyl cytosine
aCGH	Array Comparative Genomic Hybridisation
AJCC	American Joint Committee on Cancer
BAC	Bacterial Artificial Chromosome
bp	Base pairs
BR	Bloom-Richardson (scoring)
BSA	Bovine Serum Albumin
CARM1	Coactivator associated arginine transfer 1
CGI	CpG island
ChIA-PET	Chromatin Interaction Analysis by Paired End Tag sequencing
ChIP	Chromatin Immunoprecipitation
ChIP-chip	ChIP on microarrays
ChIP-seq	ChIP sequencing
chr	Chromosome
CTCF	CCCTC-binding factor
d^2	Interphase separation squared
d^2/r^2	Interphase separation divided by nuclear area
DAPI	4,6-diamino-phenylindole
DCIS	Ductal Carcinoma In Situ
dH20	distilled water
DMEM	Dubleco's Modified Eagle's Serum
DMR	Differentially Methylated Region
DMSO	Dimethyl sulpoxide
DNA	Deoxyribonucleic acid
DNMT	DNA methyl transferase

dNTP	Deoxyribonucleotides
DTT	Dithiothreitol
ENCODE	Encyclopaedia of DNA Elements
ER	Estrogen Receptor
ES	Embryonic Stem (cells)
EtOH	Ethanol
FCS	Fetal Calf Serum
FISH	Fluorescence *In Situ* Hybridisation
FITC	Fluorescein isothiocyanate
g	Gram
GEO	Gene Expression Omnibus
GTE	Glucose Tris EDTA
H3K27me3	Trimethylation of histone H3 at Lysine 27
HCl	Hydrochloric acid
HDAC	Histone deacetylase
HE	Hematoxylin-eosin
HMT	Histone methyl transferases
hr	Hours
IBC	Inflammatory Breast Cancer
IDC	Invasive Ductal Carcinoma
IHC	Immunohistochemical (staining)
ILC	Invasive Lobular Carcinoma
kb	Kilobase
l	Litre
LB	Luria-Bertani
LREA	Long Range Epigenetic Activation
LRES	Long Range Epigenetic Silencing
M	Molar
MAA	3:1 Methanol:Acetic acid
Mb	Megabase
mg	Milligram
min	Minute

ml	Millilitre
mm	Millimetre
mM	Millimolar
mmol	Millimoles
MRES	Multiple Regional Epigenetic Silencing
mRNA	messenger RNA
MW	Mann-Whitney
NaCl	Sodium chloride
NaOH	Sodium hydroxide
ng	Nanogram
nl	Nanolitre
nm	Nanometre
nM	Nanomolar
nmol	Nanomoles
NPI	Nottingham Prognostic Index
OMIM	Online Mendelian Inheritance in Man
p	'*petite*' chromosome arm
PBS	Phosphate buffered saline
PCR	Polymerase Chain Reaction
PR	Progesterone Receptor
PR	Progesterone Receptor
PRC1/2	Polycomb repressive complex 1 / 2
PRMT	Protein arginine methyl transferase
q	'*grande*' chromosome arm
qPCR	Quantitative PCR
r^2	Radius squared
RER	Region of Epigenetic Regulation
RIDGE	Region of Increased Gene Expression
RNA	Ribonucleic acid
RT	Room Temperature
RT-PCR	Reverse Transcriptase PCR
s	Seconds

SDS	Sodium Dodecyl Sulphate
SRCS	Spearman Rank Correlation Score
SSC	Saline sodium citrate
TAD	Topologically Associated Domains
TAE	Tris acetate buffer
TCM	Transcriptional Correlation Map
TCS	Transcription Correlation Score
TDLU	Terminal Duct Lobular Unit
TE	Tris EDTA
TR	Texas Red
TSA	Trichostatin A
TSS	Transcription Start Site
µg	Microgram
µl	Microlitre
µm	Micrometre
µM	Micromolar
µmol	Micromoles

Abstract

Epigenetic misregulation of gene expression is known to be an important feature in cancer. This has mainly been studied at the level of changes in DNA methylation and histone modifications at individual genes. In this thesis I have set out to investigate whether there are long-range changes in chromatin structure linked to altered gene expression in breast cancer. From large published datasets, I used a computational approach to identify large genomic regions which are coordinately misregulated in breast cancer independent of copy number aberrations (genomic effects). I found 26 regions of co-ordinate regulation of neighbouring genes that are consistent between breast tumours and breast cancer cell lines. These regions had different expression phenotypes (activation, repression, no change) compared to normal breast and also with tumour sub-type (luminal vs basal and ER status). The regions of epigenetic regulation (RER) identified in breast cancer were mostly cancer type specific.

I investigated the mechanism of long-range misregulation at one such region on chromosome 16p11.2 which is aberrantly activated in breast cancer. Interestingly, in estrogen-receptor positive (ER+ve) cells, genes in this region are upregulated relative to estrogen receptor negative (ER-ve) cells. Using fluorescence in situ hybridisation (FISH) I found that in ER+ve breast cancer cell lines and tumour tissue this region is in a more decondensed chromatin architecture than in ER-ve cell lines and tumour tissue. Furthermore this region was very compact in a normal breast epithelial cell line and breast tissue corresponding to what would be expected from the expression data. Estrogen was found to play a key role in maintaining the aberrant decondensation of chromatin at this locus on chr16p11.2, as shown by compaction of the region by starving ER+ve cells of estrogen and decompaction upon subsequent estrogen treatment. Interestingly there was also an estrogen mediated repositioning of the 16p11.2 RER domain away from the nuclear centre in hormone starved conditions and towards the centre upon estrogen stimulation. Together these results show that estrogen is key to regulating the changes in nuclear organisation and chromatin decompaction at this locus, which are associated with aberrant patterns of gene expression in ER+ve breast cancer.

Chapter 1

Introduction to Thesis

1.1 Breast Cancer

The term breast cancer describes a group of complex and molecularly heterogeneous disorders with distinct clinical and pathological differences. Cancer incidence and mortality data reveal that breast cancer is not only the most frequently occurring malignancy in women but also one of the top three causes of death due to malignancy in Europe (Ferlay et al., 2007). Within the UK >48,000 women are diagnosed with breast cancer each year and 1,000 women with breast cancer die (from www.breakthrough.org.uk). The number of women surviving breast cancer has been vastly improved by campaigns to increase awareness as well as advances in research to improve screening and treatment.

1.1.1. Normal breast anatomy & physiology

Glandular, fatty and fibrous tissues comprise the normal breast, which is situated over the pectoral muscles of the chest. Fatty tissues surround the glandular tissue containing lobules and lactiferous ducts responsible for producing milk (Figure 1.1). Each duct opens towards the nipple to form the ampulla. Connective fibrous tissue is interspersed between the lobules and ducts of the breast tissue. In the normal breast, lymph fluid is used to transport waste products around the lymphatic system. Breast cancer cells are able to spread via the lymphatic vessels and the more lymph nodes in which tumour cells are present the more likely the tumour cells are to spread to other organs as they are then able to invade the circulatory system.

During embryonic development the mammary glands are initially visible as placode-like formations from the ventral epidermis which invade the mesenchymal cell layer known as the mammary fat pad. Mammary growth is then arrested until puberty when the mammary glands undergo elongation and secondary branching to form the tubular ducts and lobules that fill the fat pad (Watson and Khaled, 2008). In the fully developed breast these lobular/ductal structures are called terminal duct lobular units (TDLU) and are embedded amongst the stroma and fatty tissue (Figure 1.1). The TDLU is composed of a bi-layer epithelium made of two distinct types of cells: the inner luminal cells that line the apical surface and the basally-positioned myoepithelial cells that surround them

(Figure 1.1 and Figure 1.2). It is thought that basal cells and breast progenitor cells also exist in the TDLU (Smalley and Ashworth, 2003; Spike et al., 2012). Furthermore it has been found that the inner luminal cells are functionally heterogenous with a hierarchy of nonclonogenic luminal cells, differentiated and undifferentiated luminal progenitors identified in normal human breast tissue (Shehata et al., 2012). In vitro mammosphere formation has been used to assay for the presence of mammary stem/progenitor cells (Dong et al., 2013; Guo et al., 2012).

During pregnancy further changes occur to the breast whereby the TDLU's undergo tertiary branching terminating in alveolar buds, accompanied by a rapid proliferation of the luminal cells and their differentiation into alveolar secretory cells (Watson and Khaled, 2008). Whilst the luminal cells are involved in milk secretion it is the myoepithelial cells with their contractile properties that are involved in transporting the milk along the ductal tree (Keymeulen et al., 2011).

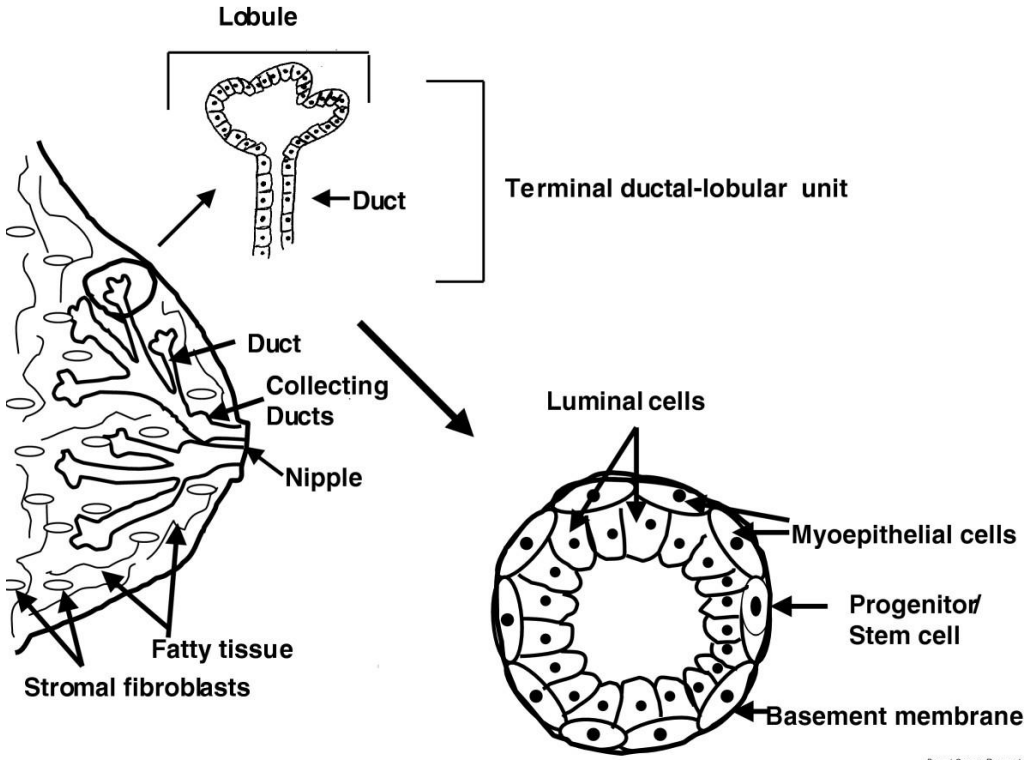


Figure 1.1: **A cross-section through the breast showing ducts and lobules.** Most breast cancer cases originate from the terminal duct-lobular unit (TDLU) which is comprised of outer contractile myoepthelial cells and inner luminal cells. (Hindle, n.d.)Image from (Dimri et al., 2005)

1.1.2. Breast cancer classification

Breast cancers derive from breast epithelial cells and can be classified according to the histopathological type, the grade of the tumour, the stage of the tumour, by the presence of specific proteins and by the gene expression profile of the tumour. This classification guides the assessment of the best possible course of clinical action and can be used as predictive factors to detect response to a specific treatment. Generally the initial classification is based on tumour histopathology - except in inflammatory breast cancer (IBC) where a physical exam is definitive of this type of ductal carcinoma by the inflammation of the breast (Giordano and Hortobagyi, 2003).

1.1.2.1. Histopathology of breast tumours

The histological grading by hematoxylin-eosin (HE) staining combined with light microscopy is the simplest method of classifying tumours of the breast from biopsies. The majority of breast cancers arise from the ducts leading to the nipples and a smaller number from the lobules where the milk is produced. Tumours arising from these areas can be described as **invasive** (abnormal growth of cells that has invaded the surrounding tissue thereby increasing the risk of metastasis to lymph nodes) or **non-invasive** (abnormal growth of cells that is localised to the site it initiated having not invaded surrounding normal tissue) (Figure 1.2).

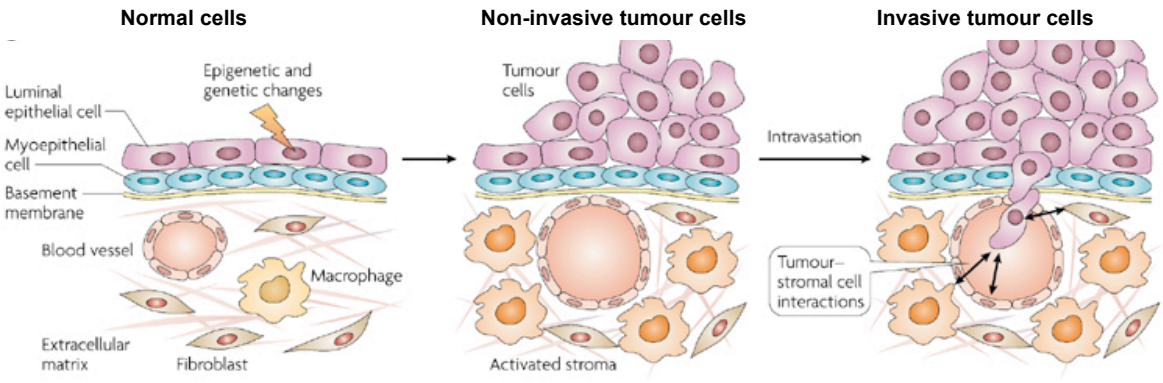


Figure 1.2: **Schematic representation of breast cancer progression.** Normal cells in the TDLUs become tumorgenic due to genetic and epigenetic lesions. These tumour cells can remain localised (non-invasive) or metastasize by invasion of the basement membrane. Adapted from (Vargo-Gogola and Rosen, 2007).

The three most common types of breast tumour are:

(1) **Invasive ductal carcinomas (IDC)** is the most common type of breast cancer representing 55% of all cases (Eheman et al., 2009). It originates in the milk duct initially spreading to the surrounding normal tissue.

(2) **Ductal Carcinoma in situ (DCIS)** represents 13% of all breast cancer cases (Eheman et al., 2009) and is the most common type of non-invasive breast cancer. It is defined as tumour cells which do not invade the basement membrane of the lactiferous ducts and is usually seen as small deposits of calcium (microcalcifications).

(3) **Invasive Lobular Carcinoma's (ILC's)** represent 5% of all cancer cases (Eheman et al., 2009) and arise from the terminal ducts of lobules. The malignant cells are generally arranged in single rows but can be difficult to distinguish from ductal carcinomas. Mostly they tend to be multicentric within a breast (approx 20% are bilateral).

1.1.2.2. Morphological grading and staging of breast tumours

As part of the pathological assessment, breast tumours are graded according to the degree of differentiation using the Bloom-Richardson (BR) scoring system (Bloom and Richardson, 1957) which assigns points dependent on whether or not each of the following features are present in a slight, moderate or marked degree:

(1) **Differentiation or tubule formation**: the tendency of the cells to group into tubular arrangement indicating the degree of structural differentiation. The higher the degree of differentiation the more it is considered indicative of good prognosis. A score of 1 indicates well-marked tubule formation, 2 - indicates moderate tubule formation and tumour sections showing little/no attempt at differentiation with cells growing in sheets/strands are scored at 3 points.

(2) **Pleomorphism**: As assessed by uniformity of the nuclear size, shape and staining. The more irregular the nuclear grade the worse the prognosis. A score of 1 indicates a fairly uniform nucleus, 2 indicates there is variation and a score of 3 that there is a severe degree of pleomorphism.

(3) **Hyperchromatic and mitotic nuclei:** A score of 1 indicates the presence of occasional hyperchromatic or mitotic figures, a score of 2 the presence of two to three and a score of 3 is given for more than this and indicates a poor prognosis. Hyperchromatic staining of the nucleus would appear to suggest grossly abnormal chromatin organisation.

Cumulative BR scores below 5 are considered Low, 6-7 is Intermediate and 8-9 is High grade. Traditionally Low scoring malignancies are known as Grade I tumours. They are usually slow growing, non-aggressive tumours with the best prognosis., High scoring Grade III tumours appear distinctly abnormal by light microscopy, are extremely aggressive, grow rapidly and tend to have the worst prognosis (Bloom and Richardson, 1957) (Figure 1.3).

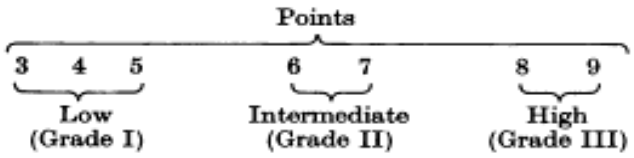


Figure 1.3: **Bloom-Richardson (BR) Grading of Tumours.** Grading is achieved by awarding points (1-3) for key features of malignant cells as viewed by light microscopy. The resultant combined score determines the grade of the tumour and is indicative of the prognosis for the patient. (taken from Bloom & Richardson., 1957).

Other systems of staging breast tumours include the American Joint Committee on Cancer (AJCC) **TNM system** that assesses the primary tumour size (T), lymph node involvement (N), and metastases (M) (Greene et al., 2002); and the **Nottingham Prognostic Index** (NPI) which adapts the BR system to include tumour size and node status (Haybittle et al., 1982).

Although histological grading is of prognostic value, much of tumour grading relies on subjective analysis of nuclear morphology, which are not always reproducible (Frierson et al., 1995). There have been more objective attempts to score nuclear morphology and texture by computational image-analysis though these are not currently widely used in clinical practice as the changes that occur in the cancer nucleus are not well characterised (Gil et al., 2002; Kiyuna et al., 2008).

1.1.2.4. Receptor status

Gross morphological grading of breast tumours is now being complemented by classification of breast tumours dependent on the expression of specific molecular markers. Breast tumours are routinely screened by immunohistochemical (IHC) staining for expression of the estrogen receptor (ER), progesterone receptor (PR) and human epidermal receptor 2 (HER2/ERBB2/neu).

As discussed later in this thesis estrogen action is mediated via the estrogen receptors alpha (ERα) and beta (ERβ) (see section 1.2.4.6.). Unless otherwise stated, throughout this thesis ER shall refer to the ERα isoform, which has been well studied in breast cancer.

ER positive (ER+ve) breast tumour cells differentiate and grow in the presence of estrogen and have a better prognosis (Nilsson and Gustafsson, 2002; Sørlie et al., 2001). Increased lifetime exposure to estrogen is one of the most important risk factors for breast cancer and drugs which can inhibit the synthesis of estrogen, or ER signalling attenuate disease progression (see section 1.1.4.3). The progesterone receptor (PR) can be used as a surrogate marker of ER activity and therefore as a prognostic factor (Mohsin et al., 2004).

ERBB2/HER2 encodes a tyrosine-kinase transmembrane receptor whose activity is associated with cell differentiation and growth (Olayioye, 2001). It is overexpressed in 20-40% of breast cancers and amplified in 20% of invasive breast tumours (Clark and McGuire, 1991; Paik et al., 1990). Generally, *ERBB2/HER2* is amplified in the more aggressive types of breast cancer, in particular inflammatory breast cancer, and is associated with very poor prognosis, increased metastatic properties and invasive potential (Sawaki et al., 2006; Tan et al., 1997).

As well as IHC, HER2 status is also assayed by fluorescence in situ hybridisation (FISH) where a FISH ratio (HER2 gene signals to chromosome 17 signals) of > 2.2 are required to classify the tumour as HER2+ve (Bertos and Park, 2011; Wolff et al., 2007).

Breast tumours with none of these receptors (ER-ve/PR-ve/HER2-ve) are known as triple negative tumours and are aggressive with the worst prognosis.

1.1.2.5. Molecular Profiling of breast tumours

Breast cancers are heterogenous in nature both between and within tumours. This has presented the biggest challenge towards studying and treating breast tumours which have been historically classified by histopathology, staging, grading and receptor status. With the development of high throughput expression profiling it is now possible to classify tumours based on their gene expression profile (Chung et al., 2002; Ramaswamy and Perou, 2003; Sørlie et al., 2003, 2001) and this disease stratification is having a large impact on the understanding and treatment of breast tumours.

Seminal work by Sorlie et al (2001), using hierarchical clustering of cDNA microarray data for 85 experimental samples including 78 breast tumours (71 ductal, five lobular and two ductal carcinomas *in situ*) led to the classification of distinct tumour subtypes based on their expression profiles (Figure 1.4) (Chung et al., 2002; Sørlie et al., 2001). The largest two groups are ER +ve and ER -ve, with at least three subtypes within the ER +ve group and three in the ER -ve group (one with normal-like gene expression, one with high HER2 gene expression and the last with basal epithelial cell gene expression features) (Sørlie et al., 2001).

Basal-type tumours highly express keratins 5 and 17, laminin, and fatty acid binding protein 7 (Figure 1.4A, red & Figure 1.4E) whereas the ERBB2+ (HER2) type tumours had highly expressed genes from the ERBB2 amplicon on chr17q22.4 (Figure 1.4A, pink & Figure 1.4C) (Sørlie et al., 2001). The normal-like tumours showed high expression of genes normally expressed in adipose and non-epithelial cell types (Figure 1.4A, green & Figure 1.4F) (Sørlie et al., 2001). This entire branch of tumours (basal-type, ERBB+ type and normal-like type) all showed a relatively higher expression of genes characteristic of basal epithelial cells, than of luminal epithelial cells (Sørlie et al., 2001).

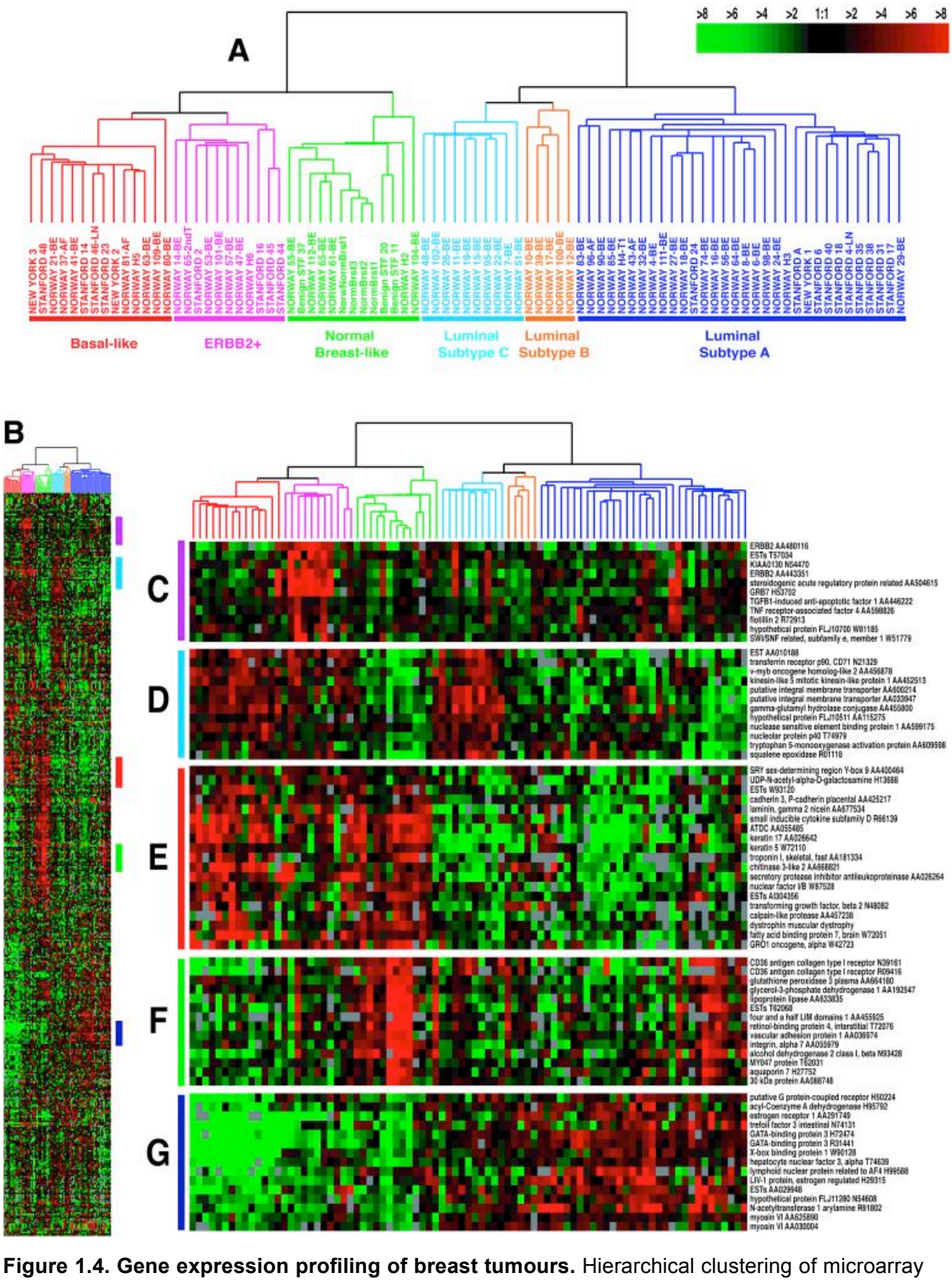


Figure 1.4. Gene expression profiling of breast tumours. Hierarchical clustering of microarray data for 85 experimental samples (including 78 carcinomas, three benign tumors, and four normal tissues) reveals distinct subtypes of tumour (***A***) Division of tumours into six subtypes based on differences in gene expression. Luminal subtype A (dark blue); luminal subtype B (yellow); luminal subtype C (light blue); normal breast-like (green); basal-like (red); and ERBB2+ (pink). (***B***) Full cluster diagram with heat map of gene expression (scaled down). The coloured bars on the right represent the heat map segments in ***C***–***G***. (***C***) *ERBB2* amplicon cluster. (***D***) Novel

unknown cluster. (***E***) Basal epithelial cell-enriched cluster. (***F***) Normal breast-like cluster. (***G***) Luminal epithelial gene cluster containing ***ER***. (figure taken from Sørlie et al 2001, PNAS).

Luminal A type tumours showed the highest expression of ERα, GATA binding protein 3, X-box binding protein 1, trefoil factor 3 (TFF3), hepatocyte nuclear factor 3 α (HNF3a), and estrogen-regulated *LIV*-1 (Figure 1.4A, dark blue and Figure 1.4G) (Sørlie et al., 2001). Remaining ER +ve tumours which also express luminal-type genes can be subdivided into luminal types B (Figure 1.4A, yellow) and C (Figure 1.4A, light blue) (Sørlie et al., 2001). Both these two groups showed low to moderate expression of luminal-specific genes (including the ER) and luminal C type tumours also have a group of highly expressed novel genes of unknown function (as did the ERBB2+ type of tumours) (Sørlie et al., 2001).

These molecular tumour subtypes were further investigated by examining their correlation with available clinical outcome data. Kaplain-Meier curves showed significant differences between the subtypes defined in Figure 1.4 for both overall survival and relapse-free patient survival (Figure 1.5) with the basal and ERBB2+ subtypes having the worst prognosis (Sørlie et al., 2001). As previously discussed, over-expression of ERBB2 (HER2) has been associated with poor prognosis and this gene was also highly expressed in the ERRB2+ tumour subtype of tumours defined by Sørlie et al (2001). Similarly the basal type tumours showed poor survival and have been associated with *TP53* mutations (Sørlie et al., 2001). *TP53* encodes p53, an important tumour suppressor, which is most frequently inactivated by mutation in many cancers and is associated with poor clinical outcome in breast cancer (Pharoah et al., 1999; Vousden and Prives, 2005) (see section 1.1.3.1 on genetic risk factors).

Overall, ER +ve luminal type tumours showed the best prognosis as highlighted by the luminal A type (highest ER expression) having the best overall survival as well as relapse free survival (Figure 1.5;(Sørlie et al., 2001). Survival analysis also showed a highly significant difference between luminal A and luminal B+C type tumours (Figure 1.5) and it has been suggested that the luminal B+C type tumours represent a group that are clinically distinct with the worst survival especially with regard to relapse (Sørlie et al., 2001).

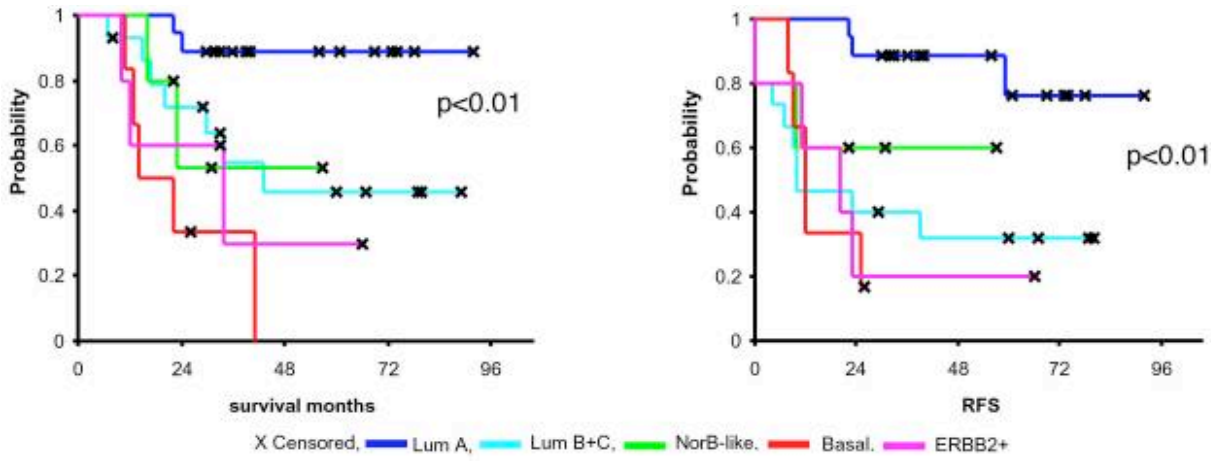


Figure 1.5. Prognosis varies between molecular subtypes of breast tumours. Probability of overall survival (left panel) and relapse-free survival (right panel) analysis of the 49 breast cancer patients based on different gene expression classification by Sørlie et al 2001. (Taken from Sorlie et al 2001).

This original molecular tumour classification was further refined in a larger number of tumours as well as independent breast tumour datasets (Sørlie et al., 2003). Clustering of 115 tumours and 7 non-malignant tissues (including the 85 used in the previous study) led to the re-classification of five distinct tumour profiles: luminal A, luminal B, ERBB2+, basal-like and normal-like ((Sørlie et al., 2003). The biggest distinction was the branching off for luminal A type tumours which were highly ER expressing compared to the other subtypes (Figure 1.6).

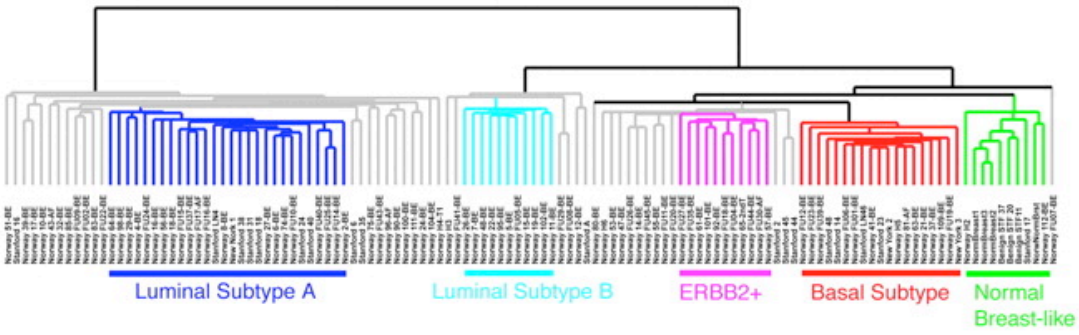


Figure 1.6: Refining the molecular subtypes of breast tumours. Dendogram showing hierarchical clustering of 115 tumour tissues and 7 nonmalignant tissues into five subgroups based on their "gene signature". Branches for tumours with low correlation to a subtype are shown in gray. (Taken from Sørlie et al., 2003)

Survival analysis showed once again that luminal A type tumours had the best clinical prognosis, and the basal and ERBB2+ subtypes the worst (Figure 1.6; (Sørlie et al., 2003). It also became apparent that BRCA1-associated tumours fell into the basal subtype and were associated with the worst prognosis and lowest expression group for ER (*ESR1* gene) and HER2 (*ERRB2* gene) (Sørlie et al., 2003). BRCA1 and BRCA2 are involved in repair of DNA damage and mutations in these genes are associated with severely increased risk of breast cancer (Fackenthal and Olopade, 2007) (see section 1.1.3.4.).

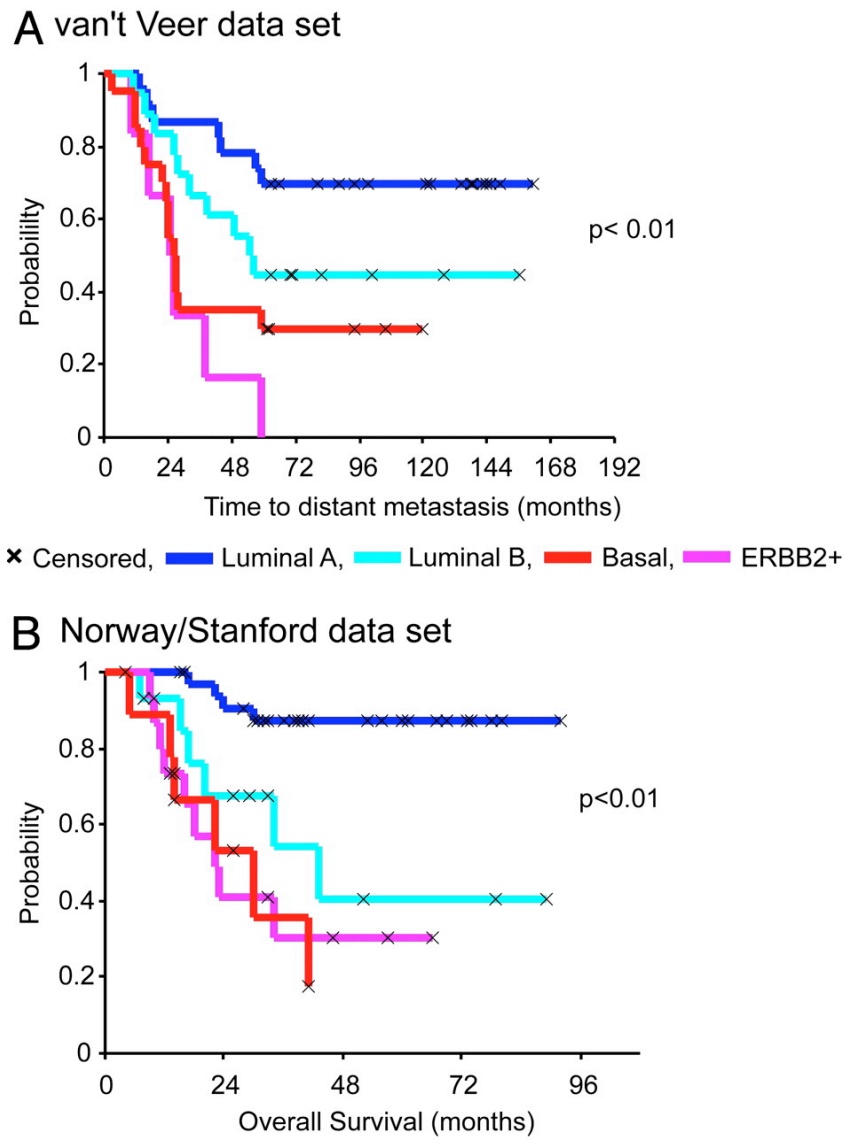


Figure 1.7. Kaplan Meier analysis of prognosis in molecular tumour subtypes is confirmed in independent datasets. (A) Time to distant metastasis development (months) in 97 cases from Veer et al data (B) Overall survival for 72 cases in an independent cohort (Norway). (Figure taken from Sørlie et al 2003).

Breast cancer cell lines have also been subject to high-throughput expression profiling and hierarchical clustering to investigate whether they reflect the molecular portraits painted in tumours (Richard M. Neve et al., 2006). Breast cancer cell lines do indeed also cluster into basal-like (ER -ve, CAV-1 positive) and luminal-like (ER +ve, ERRB3 +ve) expression patterns recapitulating overall two of the major groups observed in tumours (Richard M. Neve et al., 2006). The luminal subgroup was homogenous whereas the basal-like subgroup consisted of a division into basal A and basal B type cell lines (Richard M. Neve et al., 2006). Basal A is thought to recapitulate the gene expression profile described for the basal tumour subtype (Chung et al., 2002; Sørlie et al., 2003, 2001). Overall it is suggested that the expression profiling of these 51 cell lines successfully portray the same molecular heterogeneity seen in tumours (Richard M. Neve et al., 2006).

Agglomerative analysis of data collected from 12 different studies for 10,159 invasive breast cancer cases confirmed the existence of these molecular subtypes defined by their gene expression signature and that these can be used as clinical markers of prognosis (Blows et al., 2010). Furthermore, the ER-negative basal subclass of tumours has been further refined into at least four main subtypes with heterogeneous clinical outcome data (Teschendorff et al., 2007).

A new molecular subtype of breast cancer - denoted claudin-low - has been more recently identified and is characterised by low/no expression of tight junction proteins claudin 3, 4 and 7 and E-cadherin (Herschkowitz et al., 2007; Prat et al., 2010). This subtype is distinct from previously identified molecular subtypes as the tumours have very low levels of HER2 and luminal markers (e.g. ER, PR), a high enrichment of epithelial-to-mesenchymal (EMT) transition markers, high expression of immune response genes (e.g. CD79b, CD14, Vav1), angiogenesis (e.g. vascular endothelial growth factor C,), cell communication, differentiation (e.g. KLF2, IL6) and migration genes (e.g. integrin a5, moesin) (Prat et al., 2010). It has been suggested that these tumours have cancer stem cell -like properties as they highly express ALDH1, a marker of breast stem cells, compared to other tumour subtypes (Ginestier et al., 2007; Prat et al., 2010).

Most triple negative (ER-ve/PR-ve/HER2-ve) tumours with the worst prognosis are basal-like or claudin-low (Prat et al., 2010). By Kaplan-Meier survival analysis the claudin-low tumours are intermediate between luminal and basal-like tumours (Prat et al., 2010) (Lu et al., 2012) Moreover, patients with claudin-low basal-like tumours show worse recurrence-free survival than those with claudin-high tumours indicating that the claudin-low status is a strong prognostic factor (Lu et al., 2012).

1.1.2.6. Genome instability in breast tumours

1.1.2.6.1 Genomic amplifications and deletions

Genomic instability is a common feature of nearly all tumours and regions of genetic gain or loss are common in breast cancer. Comparative genomic hybridisation (CGH), using tumour DNA and normal reference DNA, in combination with array technology has allowed breast tumour amplifications and deletions to be assessed genome-wide (Chin et al., 2006; Fridlyand et al., 2006; Jönsson et al., 2010; Richard M. Neve et al., 2006). One study looking at the relationship between expression and genomic profiling classified breast tumours into three distinct categories (1a/16q, amplifier and complex subtype) based on their copy number profiles and this correlated with prognosis (Chin et al., 2006). Remarkably, breast cancer cell lines mirror the recurrent copy number aberrations frequently found in primary tumours, indicating that they have maintained expression and genomic patterns characteristic of the breast tumours from which they were derived (Richard M. Neve et al., 2006).

Higher resolution CGH analysis of 359 tumours revealed a detailed map of 31 frequent copy number alterations (Figure 1.8) with the most frequent aberrations found in chromosomes 1q, 8p, 8q, 11q and 16q (Jönsson et al., 2010). Overall 6 genomic subtypes of breast cancer with different clinical outcome were identified by this study (17q12, basal-complex, luminal-simple, luminal-complex, amplifier, and mixed subtype) that were similar to previously identified subtypes based on gene expression signatures (Jönsson et al., 2010).

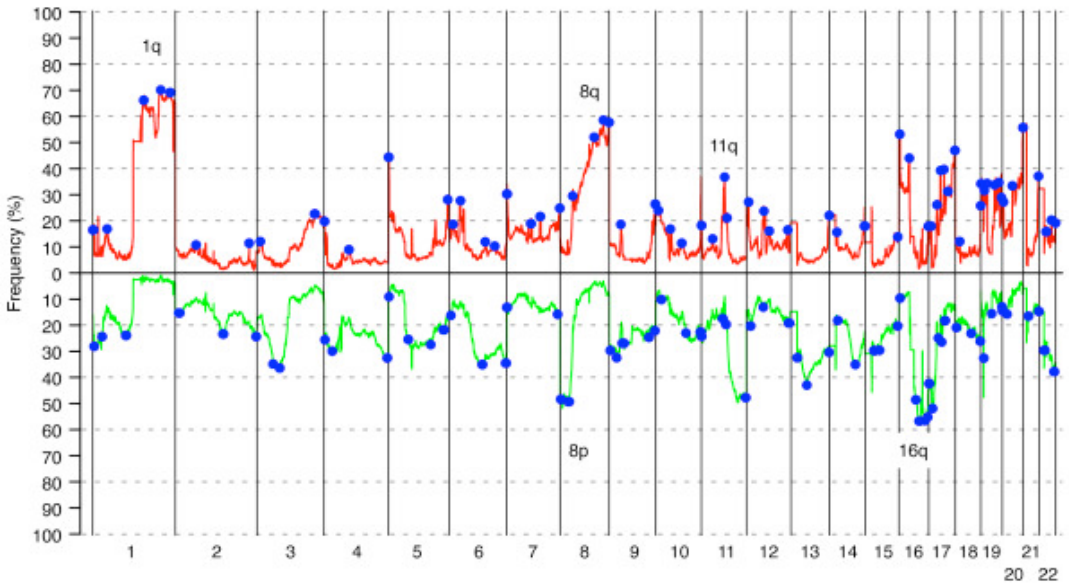


Figure 1.8. Copy- number aberrations breast cancer. Frequency plot based on aCGH data for 359 tumours where blue indicates a region of significant genomic alteration, red indicates a region of genetic gain and green a region of genetic loss. (Figure taken from (Jönsson et al., 2010)).

Integration of CGH profiles, single nucleotide polymorphisms (SNPs) and transcription profiles for 2,000 breast tumours, representing all tumour types, has allowed further refinement of these subtypes and the elucidation of 45 regions of misregulation of genes by amplification or deletion sequence changes (Curtis et al., 2012). It was therefore possible to identify 10 novel subtypes of breast cancer that were associated with different clinical prognoses (Curtis et al., 2012).

1.1.2.6.2 Deep sequencing and integrative genome analysis

The advent of array technology allowed the development of molecular signatures in tumours through analysis of differential gene expression and copy number profiles. The evolution of deep sequencing and integrative genome analysis means that there are increasing numbers of studies working towards characterising these changes on a genome-wide level and investigating the underlying mechanism(s). For example genome-wide integrative analysis of triple negative breast tumour/matched-normal tissue sequencing data has revealed a number of loci with monoallelic expression due to loss of heterozygosity, resulting in the disruption of key pathways such as the cell cycle (Ha et

al., 2012). A number of other studies have also highlighted the genetic heterogeneity of breast cancer both within and between different subtypes:

Whole-genome sequencing of 100 breast tumours has revealed strong mutational heterogeneity with 40 genes identified containing either copy number changes or/and driver point mutations (Stephens et al., 2012). The incidence of these mutations was highly variable with a maximum of 6 tumourigenic mutations per sample. Moreover some of the mutations were shown to be associated with age in ER-ve tumour samples but not ER+ve patients (Stephens et al., 2012). Deep-sequencing and transcriptome analysis of 104 primary triple-negative breast cancer genomes revealed the presence of mutations and copy number aberrations which fluctuate greatly between (and within) tumour samples, indicating that these changes occur during different stages of tumour progression (Shah et al., 2012). Three of the genes identified (*TP53*, *PIK3CA* and *PTEN*) were associated with early breast tumour development (Shah et al., 2012).

Exome sequencing of 103 tumour/matched-normal paired samples from patients revealed a number of heterogeneous subtypes due to somatic non-silent mutations (Banerji et al., 2012). This study also implicated the *MAGI3–AKT3* translocation and mutations in the transcription factor *CBFB* and its partner *RUNX1* in breast cancer for the first time (Banerji et al., 2012). The *MAGI3-AKT3* gene fusion was specific to triple negative breast cancer and caused constitutive activation of AKT kinase (Banerji et al., 2012).

The mutational heterogeneity of breast tumours was also highlighted by whole-genome sequencing of the ER+ve breast cancer genome in patients undergoing treatment with aromatase inhibitors (see section 1.1.4.3) (Ellis et al., 2012). Patients with high expression of Ki67, a marker of proliferation and implicated in resistance, had a higher frequency of somatic mutations compared to those with low Ki67 expression (Ellis et al., 2012).

Integrative analysis of six different technology platforms, including genomic copy number arrays and exome sequencing, did not reveal any previously unknown molecular subgroups based on expression profiling (as originally defined by Sørlie et al., 2003 and confirmed in subsequent studies by Blows et at., 2010 amongst others). However the

mutational heterogeneity of the breast cancer genome was once again highlighted with numerous subtype-specific novel gene mutations identified (Koboldt et al., 2012).

1.1.3. Breast cancer risk factors

The exact etiology of breast tumourignesis is unknown but a number of factors can increase the risk of developing breast cancer. These risk factors can be biological, genetic or environmental. Age is the most significant risk factor after gender. Lifetime risk of developing breast tumours increases significantly with age, with the estimated risk for a 29-year of 1 in 2000, compared with 1:50 for a 49-year old and 1:13 at age 69 (Sasieni et al., 2011). Childbearing and breast feeding are associated with reduced risk of developing breast cancer, with a 4.3% decrease in risk for every 12 months spent breast feeding and 7% decrease for every childbirth ("Breast cancer and breastfeeding," 2002). Hormone replacement therapy (HRT) for five or more years increases the risk of developing breast cancer by 35% but this disappears 5 years after stopping HRT use (Collaborative Group on Hormonal Factors in Breast Cancer, ("Breast cancer and hormone replacement therapy," 1997). Increased risk has been observed in both estrogen-only as well as estrogen-progestin therapy with highest risk for the latter (Stahlberg et al., 2004).

A high breast density (ratio of tissue to fat) has been highly correlated with development of tumours in a number of population studies and meta-analyses making it one of the strongest risk factors for breast cancer (Boyd et al., 2010; McCormack and dos Santos Silva, 2006).

Breast cancer risk increases by 11 in 1000 per additional alcoholic beverage per day (Allen et al., 2009). Many epidemiological studies have also shown a dosage dependent increase in the risk of developing breast cancer upon alcohol intake, the mechanism of which is unknown, although alcohol consumption increases estrogen levels which is also an important risk factor (reviewed in (Coronado et al., 2011). Low fat diets are thought to decrease risk of breast cancer development as well as reducing the risk of recurrence (Chlebowski et al., 2006).

1.1.3.1. Genetic risk factors

Constitutional mutations in key genes required for normal cell function, especially the response to DNA damage, cell signalling and cell adhesion, can predispose an individual to developing cancer. These mutations show autosomal dominant inheritance but incomplete penetrance. According to Knudson's two-hit hypothesis this is because although there is an inherited mutation (first "hit"), a secondary somatic mutation (second "hit") is required in the target tissue for tumourigenesis to occur due to complete loss of function. This can occur through a second independent mutation or due to loss of heterozygosity. Genetic testing can be used to screen individuals for risk mutations (first "hit") especially where there is a family history of breast cancer.

BRCA1* and *BRCA2: Hereditary breast cancer accounts for 5-10% of breast cancer cases in Europe and ~80-90% of these are due to mutations in *BRCA1* and *BRCA2* (Ford et al., 1998; Gage et al., 2012). Lifetime risk for carriers of mutations in *BRCA1* and *BRCA2* genes is 50-80% (Fackenthal and Olopade, 2007). Tumours with triple negative receptor status tend occur in *BRCA1* mutation carriers rather than *BRCA2* and tend to be higher grade with the worst prognosis (Mavaddat et al., 2012; Reis-Filho and Tutt, 2008). Receptor status amongst *BRCA1/2* tumours is highly variable with ER+ve/PR-ve tumours less likely to be *BRCA2* mutated than ER+ve/PR+ve tumours, and ER-ve/PR+ve tumours were more likely to be *BRCA2* mutated than ER-ve/PR-ve tumours (Mavaddat et al., 2012). BRCA1 and BRCA2 are both involved in repair of double-stranded breaks in DNA by homologous recombination (HR) where they function at different stages (Figure 1.9). BRCA1 is involved at the early stages when it is phosphorylated in response to DNA damage and becomes involved in checkpoint activation of signalling pathways (Turner et al., 2005). BRCA1 is a RING finger protein that associates with another RING domain protein BARD to form a heterodimeric E3 ubiquitin ligase. C-terminal BRCT repeats can bind phosphorylated proteins and it is this function, but not E3 ligase activity, that is required for tumour suppression (Shakya et al., 2011). BRCA2 mediates HR by binding to, stimulating the oligomerisation of, and localizing the recombinase RAD51 to sites of DNA damage (Turner et al., 2005).

ATM encodes a kinase involved in DNA damage repair that initiates a signalling cascade which includes phosphorylation of p53, BRCA1 and CHK2 (Figure 1.9). Autosomal recessive mutations in ATM cause Ataxia telangiectasia which is characterised by immune problems. Mutations in *ATM* are also more prevalent in families with a history of breast cancer (Thorstenson et al., 2003). There is a moderate increase in the risk for developing breast cancer in individuals with heterozygous mutations in *ATM* (Thompson et al., 2005).

TP53: This tumour suppressor gene encodes a p53 which has many cellular roles including activation of DNA repair, cell cycle arrest and apoptosis. Constitutional mutations in *TP53* are responsible for Li Fraumeni Syndrome and breast cancer accounts for 24% of the cancer incidence in these families (Gage et al., 2012; Kleihues et al., 1997). There is a high incidence of *TP53* mutations in BRCA1 breast tumour cases compared to non-BRCA1 tumours (Holstege et al., 2009).

CHEK2: This cell cycle checkpoint kinase is activated by ATM and is involved in the DNA damage response (Figure 1.9) (Matsuoka et al., 1998)**.** Premenopausal women with breast cancer with *CHEK2* mutations have been shown to be at high risk of tumour recurrence and poorer prognosis compared with non-carriers (Schmidt et al., 2007). *CHEK2* mutations are also present in families with Li-Fraumeni Syndrome (Bell et al., 1999).

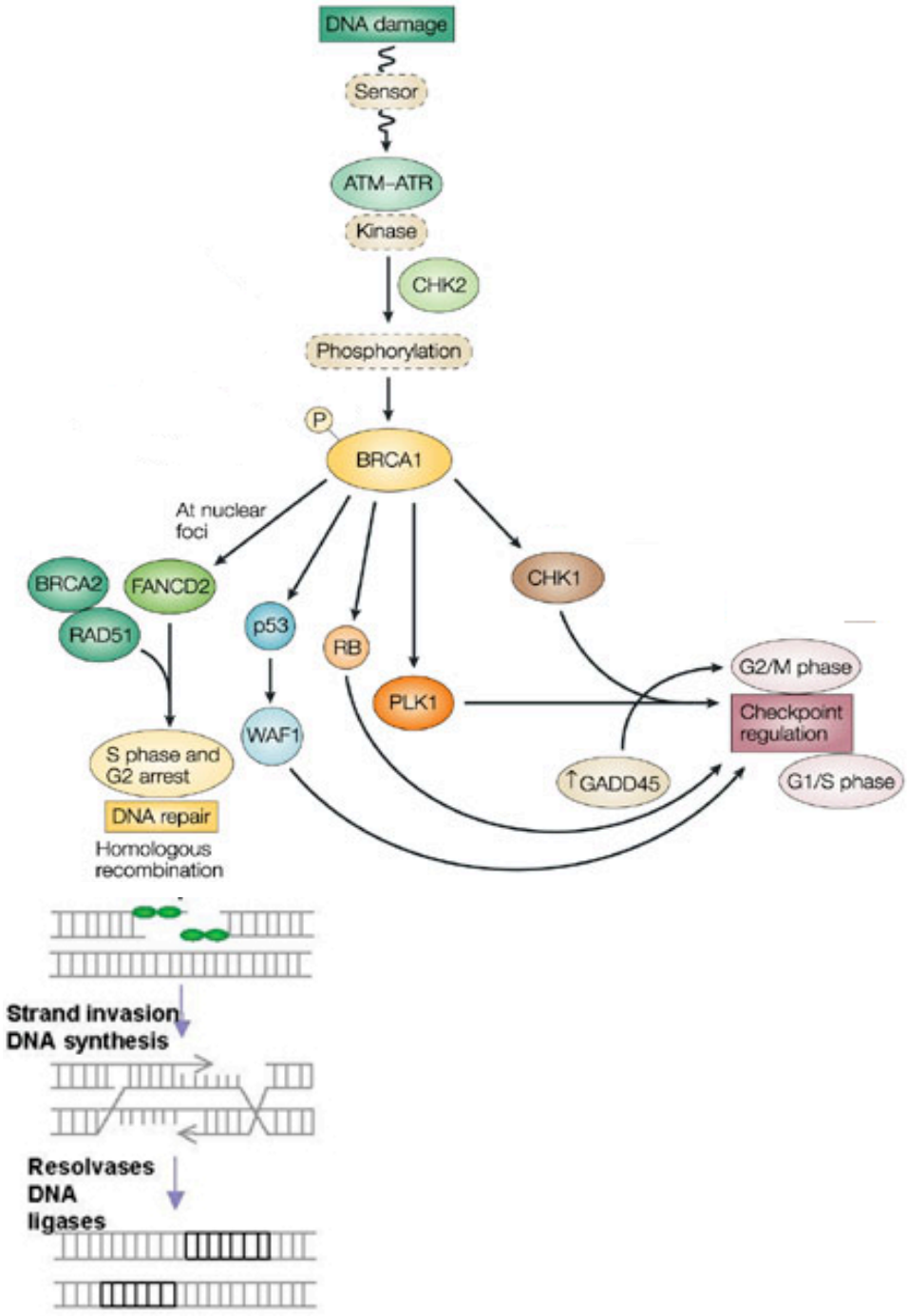


Figure 1.9. The role of breast cancer risk genes in the DNA damage response. (A) In response to DNA damage in normal cells, ATM phosphorylates BRCA1, leading to formation of a complex with BRCA2 which relocates to DNA damage sites to initiate repair by homologous recombination. (Adapted from (Mostoslavsky, 2008; Narod and Foulkes, 2004)

PTEN: Mutations in *PTEN* cause Cowden & Bannayan-Zonana syndromes which, though rare, have been associated with severely increased breast cancer risk (Marsh et al., 1998; Schrager et al., 1998). *PTEN* encodes a lipid phosphatase involved in inhibiting the PI3K/Akt/mtor signalling pathway which regulates proliferation, apoptosis and metabolism (Figure 1.10). Mutations in *PTEN* cause hyper-activation of the PI3K/Akt/mtor signaling pathway in breast cancer (and other malignancies) where is contributes to survival of tumour cells and these patients suffer from poor prognosis (Ghayad and Cohen, 2010).

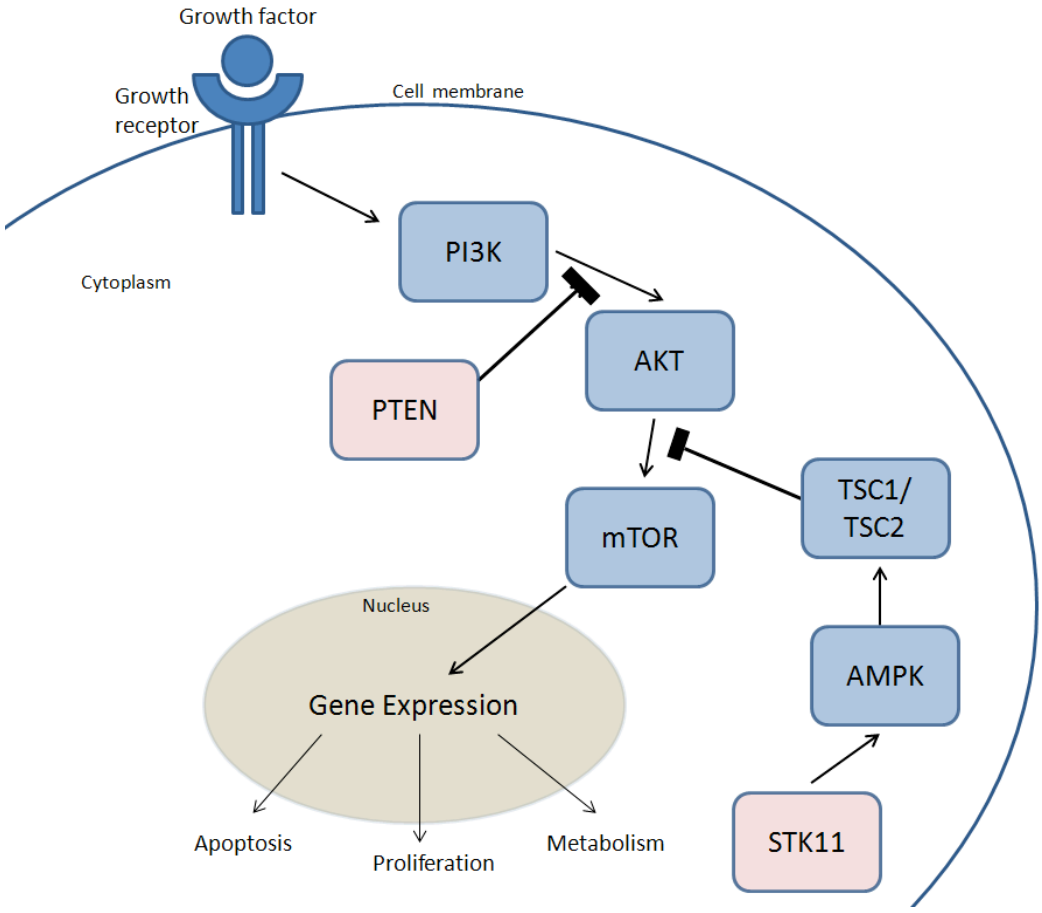


Figure 1.10. Schematic representation of the PI3K/Akt/mTOR signalling pathway. PTEN functions to inhibit the PI3K/AKT pathway progression preventing activation of AKT and mTOR which would result in upregulation of cellular processes which are important to tumourigenesis. STK11 mutations inhibit signalling through the AMPK pathway therefore preventing activation of mTOR by the TSC1/2 complex.

STK11: Patients with Peutz-Jeghers syndrome, characterised by oral pigmentation and gastrointestinal polyps, have mutations in *STK11* and increased risk of developing breast cancer (Hearle et al., 2006). *STK11* encodes a serine/threonine kinase that – like PTEN - is involved in PI3K/Akt/mtor pathway inhibition (Figure 1.10).

CDH1 encodes E-cadherin which is important for cell-to-cell adhesion and is commonly mutated in hereditary diffuse gastric cancer (Becker et al., 1994). Patients with familial *CDH1* mutations are highly predisposed to developing invasive lobular breast cancer (Schrader et al., 2008).

A number of studies based on consortium data have led to the discovery of low penetrance susceptibility loci with heterogenous associations with breast cancer subtype (Broeks et al., 2011; Garcia-Closas and Chanock, 2008; Garcia-Closas et al., 2008; Reeves et al., 2010). Recently an investigation of genome-wide association study (GWAS) susceptibility loci in a pool of 31 studies showed that 6/8 were associated with ER+ve tumours; 4/8 with triple negative tumours; and two loci, ***CASP8*** and ***TGFB1***, were high risk loci for PR-ve tumours (Broeks et al., 2011). This heterogeneity was similar to previous findings using GWAS risk loci in ***FGFR2, TNRC9, MAP3K1***, **8q24** and ***LSP1***, which showed heterogeneity with regard to ER status (Garcia-Closas et al., 2008).

1.1.4. Current therapies for breast cancer

Multiple therapeutic strategies exist for breast cancer depending on the tumour classification. This is an evolving field with increasingly targeted therapies thanks to the advent of molecular profiling/genetic markers and new models for predicting response to the different treatment options. It is an exemplar for the advantages of stratified medicine.

1.1.4.1. Surgery and radiotherapy

Removal of a breast tumour by surgical excision is known as a lumpectomy, but the margin of surrounding tissue removed is hotly debated with 20-30 % of patients requiring secondary re-excision surgery (Morrow et al., 2012). Radiotherapy usually follows a lumpectomy to kill any remaining tumour cells in the breast (radioadjuvant therapy) or

prior to surgical intervention to shrink the tumour to a size suitable for excision (neoadjuvant radiotherapy) (Kirova, 2010). Another treatment option is mastectomy - removal of the whole breast by either a simple mastectomy (removal of breast tissue) or more rarely a radical mastectomy (removal of all breast tissue and underlying chest muscle). This can also be followed by radiotherapy, particularly if there has been lymph node involvement and it was not possible to remove all nodes surgically (Kirova, 2010). Individuals with strong family histories of breast cancer, in particular *BRCA1/BRCA2* mutation carriers, often undergo mastectomies as a preventative measure.

1.1.4.2. Chemotherapy

Chemotherapy with different combinations of cytotoxic drugs can be used to kill tumour cells in either a neoadjuvant setting to shrink the tumour or adjuvant setting to kill remaining tumourigenic cells. The drugs (anthracyclins, taxanes, anti-metabolites, DNA replication inhibitors, alkylating agents, topoisomerase inhibitors etc) can have a number of side effects (infections, hair loss, nausea and vomitting). Frequently used treatment combinations are CMF (cyclophosphamide, methotrexate and 5-fluorouracil), FAC/CAF (5-fluorouracil, doxorubicin and cyclophosphamide) and FEC (5-flurouracil, epirubicin and cyclophosphamide) (Klaassen and Seeber, 1997).

1.1.4.3. Endocrine therapy

Reducing the level of circulating estrogen or blocking its action using drugs is widely used to target ER +ve tumours. Selective estrogen receptor modulators (SERMs) function by competitively occupying the estrogen binding site on estrogen receptors and these include Tamoxifen, Toremifene, Raloxifene, Idoxifene and Droloxifen (Goldhirsch et al., 2002). Multiple clinical trials have provided evidence that treatment with SERMs significantly reduces the risk of developing ER+ve invasive breast cancer and that Tamoxifen can reduce the risk by 50% (Cazzaniga and Bonanni, 2012; Visvanathan et al., 2009). Alternatively, estrogen hormone synthesis can be blocked by aromatase inhibitors which preventing the conversion of androstenedione and testosterone to oestrone and oestradiol by this enzyme (Figure 1.11). Aromatase inhibitors can be steroidal (Type I) or non-steroidal (Type II) and include aminoglutethimide, anastrozole, letrozole, 4-OH-androstendione and exemestane. A number of studies have shown that

use of aromatase inhibitors can even be more effective than standard treatment with SERMs for shrinking tumours prior surgery (Dixon et al., 2000, 1999) with improved disease-free survival and fewer side effects (Cazzaniga and Bonanni, 2012). Pituitary downregulators can also be used to inhibit the release of gonadotrophin from the pituitary thereby preventing the release of estrogen from the ovaries (Goldhirsch et al., 2002; Tan and Wolff, 2007).

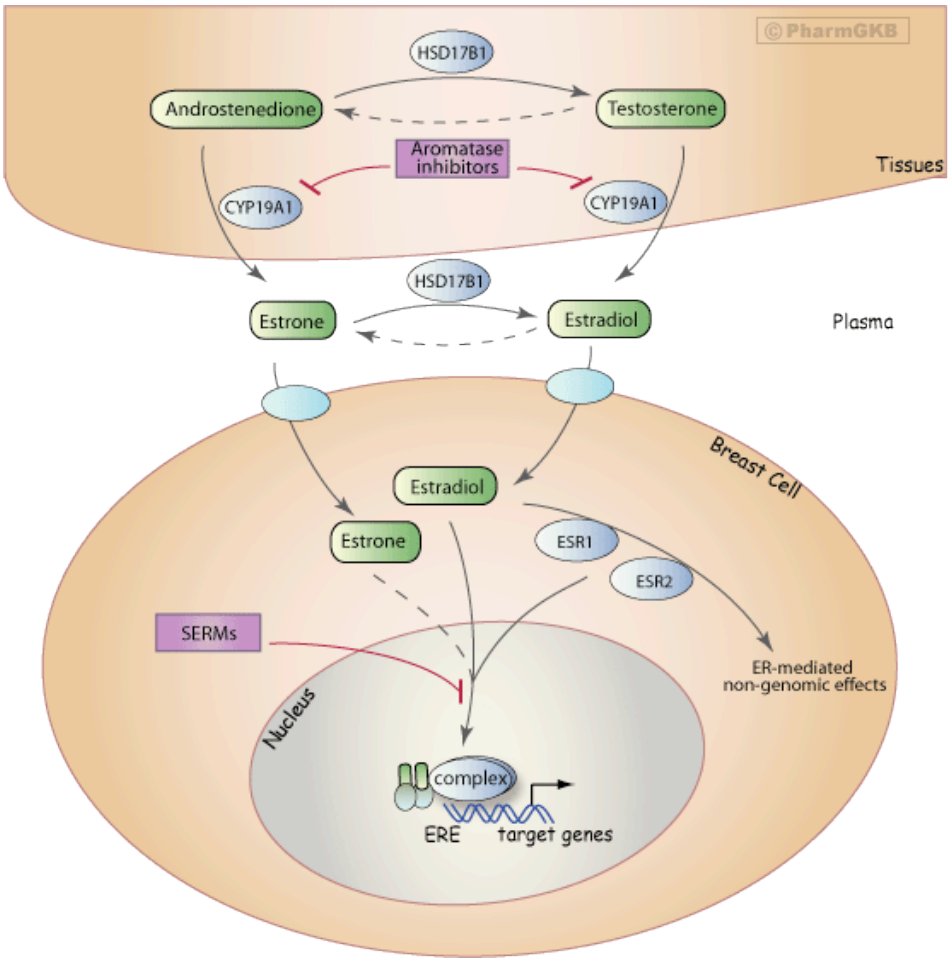


Figure 1.11. Summary of estrogen synthesis and antiestrogen therapy. Estrogen is synthesized by aromatase enzyme from androgens. Estrogen binds to the estrogen receptors (ERb and ERa), where they recruit coactivators/repressors leading to dimerization, conformational change and allow the binding of estrogen response elements (EREs) upstream of estrogen responsive genes. Endocrine therapy by treating with inhibitors of aromatase can be used to prevent estrogen synthesis or by SERMs which bind to the ERs inducing conformational changes that inhibit transcription of estrogen regulated genes. Image from PharGKB.org (Whirl-Carrillo et al., 2012) with permission from Stanford University.

1.1.4.4. Molecular targeted therapies

Tumours overexpressing HER2 can be treated with drugs which inhibit this receptor including monoclonal antibodies against HER2 such as Tratuzumab (Herceptin) and Pertuzamab (Omnitarg), or tyrosine kinase inhibitors (Lapatinib /Tyverb) (Chu and Lu, 2008). Moreover combined treatment of Lapatinib and Tratuzuab with aromatase inhibitors have proved more effective in hormone receptor positive cases of HER2 amplified metastatic tumours (Fleeman et al., 2011)

Patients with cancer due to *BRCA1/2* mutations can be treated with inhibitors against Poly(ADP-ribose) polymerase (PARP). PARP is essential for the base excision repair (BER) pathway that is responsible for repairing single-stranded breaks in DNA (Ashworth, 2008; Baute and Depicker, 2008; Rouleau et al., 2010). Inhibition of PARP prevents repair by BER leading to double-stranded breaks which then cannot be repaired by HR in *BRCA1/2* deficient cells as they are key to that pathway (see section 1.1.3.1 and figure 1.9). PARP inhibition in *BRCA1/2* mutant cells leads to an accumulation of DNA damage and cell death due to synthetic lethality (Ashworth, 2008; Carroll et al., 2011). Olaparib/AZD2281 is a PARP inhibitor that is currently undergoing phase I clinical trials (Bundred et al., 2013). Interestingly, in olaparib-resistant metastastic tumours secondary mutations in *BRCA2* have been identified which are able to restore full-length functional BRCA2 and are thought to contribute to the mechanism of resistance to this PARP inhibitor (Barber et al., 2013).

Angiogenesis is a requisite for tumour growth and therefore targeting tumour vasculature can be used to treat tumours. Vascular endothelial growth factor (VEGF) which is needed for angiogenesis can be blocked by the tyrosine kinase inhibitor Sunitinib (Sutent) or by the monoclonal antibody Bevacizumab (Avastin) (Carroll et al., 2011).

1.2. Chromatin function and organisation

The human genome comprises 3.2 billion base pairs (bp) of DNA wound around proteins and packaged into the cell's nucleus. Together the combination of DNA wound around proteins is known as chromatin. Packaging of chromatin in the nucleus is not random but infact highly organized with important roles in the regulation of gene expression.

1.2.1. DNA methylation

Methylation of DNA at CpG dinucleotides (5mC) is a heritable epigenetic mark catalysed by DNA methyltransferases (DNMT) and often associated with repression of transcription (Figure 1.12). Methylation occurs frequently in the human genome at 70-80% of CpGs (Ehrlich et al., 1982). CpG dinucleotides are not evenly distributed in the genome and cluster together to form 'CpG islands (CGIs)' which span 0.5 to 5kb and are generally unmethylated in normal cells (Weber et al., 2007). CpG islands occur at the promoters of 60-70% of all human genes (Weber et al., 2007) and in particular constitutively expressed genes (e.g. housekeeping genes) but just 40% of tissue specific genes (Larsen et al., 1992). Although most CGIs coincide with transcription start sites (TSS) of genes a number of "orphan" CGIs exist that are not associated with known promoters and undergo DNA methylation during development (Illingworth et al., 2010). In general it is thought that unmethylated CGIs are permissive of transcription due to their association with RNA polymerase II recruitment, active chromatin structure and transcription factor binding which becomes inhibited directly by DNA methylation or indirectly through recruitment of proteins and co-repressors (reviewed in (Deaton and Bird, 2011)).

Another major modification to the DNA itself is the hydroxylation of 5mCto form hydroxy-methyl-cytosine (5hmC) catalysed by ten-eleven-translocation (TET) enzymes (Figure 1.12) (Iyer et al., 2009). It has long been known that this epigenetic mark is present in bacteriophages and trypanosomes but it was only recently found to be an abundant modification in mammalian genomes (Kriaucionis and Heintz, 2009). Bisulphite sequencing, which has been traditionally used to detect DNA methylation, does not distinguish between hydroxyl-methylation and methylation of DNA (Jin et al., 2010; Nestor et al., 2010). Methods that distinguish between 5mC and 5hmC include end-

labeling with thin-layer chromatograph, enzymatic radio-labeling by glycosylation, real time sequencing, and sensitive mass spectrophotmetry (reviewed in (Kinney and Pradhan, 2013)).

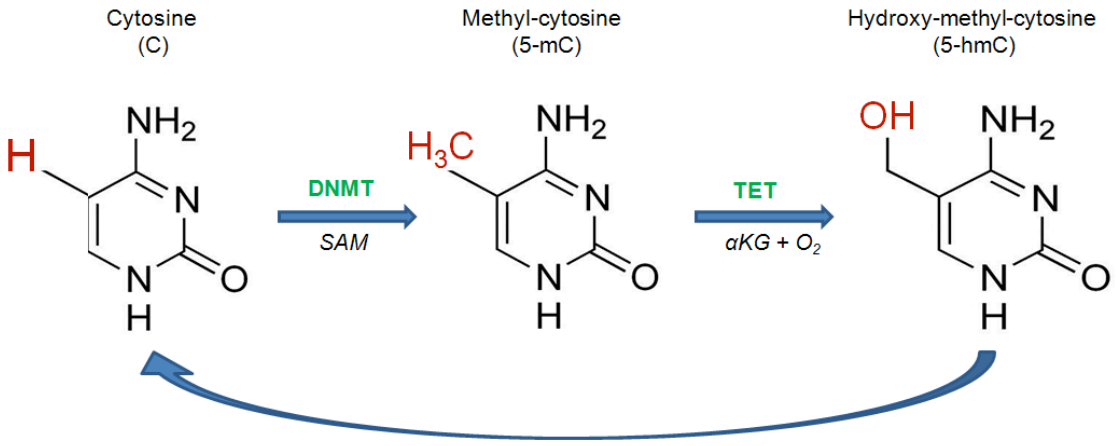


Figure 1.12: Summary of the major modifications to Cytosine at position 5. Cytosine (C) is converted to methyl-cytosine (5-mC) by DNA methyltransferase (DNMT) enzymes using the co-factor S-adenosyl methionine (SAM). Methyl-cytosine can then be converted by Ten-eleven translocation (TET) enzymes with α-keto glutarate (αKG) and oxygen to hydroxy-methyl-cytosine (5-hmC). Hydroxy-methyl-cytosine can then be converted back to Cytosine by a combination of TET enzymes and thymine DNA glycosidase (TDG).

The exact function of 5hmC is currently not known though it is known to be on the pathway to DNA demethylation. 5hmC has a much reduced affinity for binding to methyl binding proteins (MBDs) compared to 5mC (Jin et al., 2010), but their might be other proteins/protein motifs with specific affinity for 5hmC. The abundance of 5hmC amongst human tissues is highly variable compared with 5mC, with the highest levels in the brain (Li and Liu, 2011; Nestor et al., 2012) and the levels rapidly decrease as cells are cultured. Moreover 5hmC is enriched in gene bodies of active genes and the level of enrichment directly correlates with the levels of gene expression (Nestor et al., 2012).

1.2.2. Chromatin Structure & Remodeling

Histones fall into two broad classes: linker (H1 and H5) and core (H2A, H2B, H3 and H4). Approximately 146 bp of DNA is wound around 8 core histone proteins (2 of each type) to form a nucleosome with linker histones at DNA exit/entry sites (Hamiche et al., 1996; Luger et al., 1997; Zhang and Reinberg, 2001). Nucleosomes form the 10nm chromatin fiber, which is visible as "beads-on-a-string'' in low salt cellular extracts by electron microscopy (Olins and Olins, 1974). Under physiological salt conditions *in vitro* the 10nm fiber is packed sequentially into 30nm fibers with 6 nucleosomes per turn and 700nm structures which might represent heterochromatin (Hansen et al., 1989; Zhang and Reinberg, 2001). Evidence for the existence of the 30nm chromatin fiber *in vivo* has been demonstrated in starfish sperm nuclei although its presence in most interphase cell types remains very contentious (as reviewed in (Bian and Belmont, 2012; Fussner et al., 2011)).

There is a nucleosome free region (NFR) at the transcription start sites (TSS) of active genes which is thought to allow polymerase and transcriptional machinery access to the gene (Yuan et al., 2005). To allow transcription to take place, there is a constant removal of histones ahead of the polymerase and replacement after it has passed. This process is known as chromatin remodeling (or "histone exchange") and is regulated by ATP-dependent chromatin remodeling complexes and covalent modifications to histones. A number of these remodeling complexes have been implicated in the regulation of the transcription activation mediated by nuclear receptors (Belandia and Parker, 2003). The SWI/SNF complex was one of the first ATP-dependent remodelling complexes to be identified and has been implicated in the regulation of ER- and BRCA1- dependent transcription (García-Pedrero et al., 2006; Harte et al., 2010).

Nucleosome organisation can also be mediated by the FACT (facilitates chromatin transcription) complex which acts as a histone chaperone promoting the accessibility of DNA possibly by eviction of a H2A-H2B dimer (Winkler and Luger, 2011). The FACT complex is also physically and functionally associated with the PAF (polymerase associated factor) complex during transcriptional elongation by RNA polymerase II

(Squazzo et al., 2002). The PAF complex recruits the chromatin remodelling factor CHD1 where it interacts with other factors involved in transcriptional elongation including components of the FACT complex, though the specific function of CHD1 is unknown (Simic et al., 2003). PAF-1, a component of the PAF complex in humans, is also associated with modifications to histones as well as chromatin remodelling complexes (Dey et al., 2011).

1.2.3 Histone modifications

Post-translational modifications can occur on the globular domain of the histones themselves but they predominantly occur at the N- and C- terminal tails that extend from the nucleosome . Histone-modifying enzymes catalyze the addition or removal of acetylation, methylation, phosphorylation, ubiquitination and ADP-ribosylation, amongst other chemical changes. These changes can affect the charge on histone tails but also provide the binding sites for chromatin proteins that recognise specific histone modifications. Histone tail modifications can be assayed genome-wide by chromatin immunoprecipitation (ChIP) with specific antibodies combined with arrays (ChIP-chip) or high through-put sequencing (ChIP-seq) (Barski et al., 2007).

1.2.3.1. Histone Acetylation

Addition of negatively-charged acetyl groups to positively-charged lysine residues, which are abundant in histone tails, neutralises the charge on histone tails and is known to relax DNA-histone interactions making chromatin permissive to transcriptional machinery (Hebbes et al., 1988; Hong et al., 1993). Acetyl group addition is carried out by the histone acetyl transferase (HAT) enzymes and is reversed by histone deacetylases (HDACs). Acetylation of histones H3 and H4 occurs at the TSS of active genes (Bernstein et al., 2005; Pokholok et al., 2005) and acetylation ‘islands’ distant from the TSS can be used to predict the presence and activity of long-range regulatory elements (e.g. enhancers of gene expression) (Roh et al., 2007), (Visel et al., 2009), (Creyghton et al., 2010).

1.2.3.2. Histone Methylation

Histones can be mono-, di- or tri-methylated by the addition of methyl group(s) to arginine (R) or lysine (K) residues by histone methyltransferases (HMT). Demethylation is carried out by LSD1 and the jumonji (JMJ) family of proteins. Histone methylation of R residues are little studied compared to lysine methylation residues but are thought to be important for cell proliferation and differentiation (as reviewed in (Wysocka et al., 2006a). These modifications include: methylation histones of H3R2, H3R17 and H3R26 by coactivator associated arginine methyl transfer 1 (CARM1); histone H3R8 by protein arginine methyl transferase 5 (PRMT5); and histone H4R3 by either PRMT1 or PRMT2. Arginine methylation is linked to the mechanism of transcriptional activation by nuclear hormone receptors (Barrero and Malik, 2006; Ma et al., 2001; Wang et al., 2001; Xu et al., 2004, 2001) and is not just limited to methylation of histones (Ceschin et al., 2011; Le Romancer et al., 2008; Naeem et al., 2007).

Tri-methylation of H3K27 (H3K27me3) is an important repressive histone mark that is catalysed by the EZH2 (enhancer of zeste 2) component of the polycomb group repressive complex 2 (PRC2). There are two main classes of polycomb repressive complexes (PRC): the PRC1 complex (PC, PH, PSC and RING1B) which can ubiquitinate H2AK119 and the PRC2 complex (EZH2, SUZ12 and EED) which can tri-methylate H3K27 (Otte and Kwaks, 2003). H3K27me3 is associated with repression of gene expression through interactions with, or recruitment of, the PRC1 to DNA(Simon and Kingston, 2009). The PRC system plays a critical role in regulating gene expression during normal development. PRC1 has a key role in compacting chromatin *in vitro* and *in vivo* (Ragnhild Eskeland et al., 2010; Francis et al., 2004) in association with gene silencing. This is antagonised by histone acetylation – which partially reverses polycomb-mediated chromatin compaction and gene repression (R Eskeland et al., 2010). Furthermore, H3K27me3 deposition and binding of polycomb components in mouse ES cells has been linked to DNA methylation through the action of Tet1 enzyme which is thought to recruit polycomb group proteins to CpG-rich promoters in mouse ES cells (Wu et al., 2011).

H3K27me3 has also been shown to mark genes that become hypermethylated in tumour cells and it is thought that PRC2 normally maintains the mark in differentiated cell types but that in tumour cells this complex recruits DNMT enzymes for *de novo* methylation (Schlesinger et al., 2007; Widschwendter et al., 2007). This has given rise to the idea that a "stem cell chromatin signature" may exist where loss of H3K27me3 can predispose genes to acquire aberrant DNA methylation and repressive histone marks in adult cancers (Ohm et al., 2007). Genome-wide studies have indicated that there is a mutually exclusive switch between DNA methylation and H3K27me3 repressive marks (Gal-Yam et al., 2008; Hahn et al., 2008) though subsets of genes have also been identified with both these repressive marks (Coolen et al., 2010; Gal-Yam et al., 2008; Hawkins et al., 2010; Meissner et al., 2008). The latter finding is supported by the advent of Bisulphite/ChIP sequencing which showed that H3K27me3-marked histones were able to bind to both methylated and unmethylated DNA at specific regions of the genome and that are repressed in cancer (Statham et al., 2012).

Histone methylation is not uniformally associated with repression of genes. H3K4me3 is deposited by members of the COMPASS family of HMTases (Shilatifard, 2012) at the TSS of genes where it is associated with transcriptional activity (Barski et al., 2007; Milne et al., 2002; Noma K et al., 2001; Wysocka et al., 2006b; Yokoyama et al., 2004). H3K36 is methylated by the Set2 HMTase that associates with the elongating (serine 2-phosphorylated) RNA polymerase II. Whereas H3K4me3 peaks at the 5' end of genes, H3K36me3 is enriched toward the 3' end and is particularly associated with exons (Bell et al., 2008; Hon et al., 2009; Santos-Rosa et al., 2002). The presence of H3K4me3 and also di/tri-methylated H3K36 (H3K36me2/3) directly inhibits PRC2 therefore preventing methylation of H3K27 on genes that are being actively transcribed (Schmitges et al., 2011).

1.2.3.3. Histone Phosphorylation

Histone phosphorylation is associated with transcriptional regulation, cell cycle progression, chromosome condensation, DNA repair and regulation of developmental genes (Cheung et al., 2000; Cruickshank et al., 2010; Johansen and Johansen, 2006; Rogakou et al., 1998). Phosphorylation of serines 10 and 28 in histone H3 (by Aurora B

kinase) correlate with transcriptionally active loci and is associated with mitotic condensation (Nowak and Corces, 2000). Phosphorylation of H3S28 mediated by mitogen- and stress-activated kinases (MSK) leads to a displacement of polycomb protein complexes from H3K27me3 marked promoters forming a double mark (H3K27me3S28p) and leading to transcriptional activation (Gehani et al., 2010). Most recently, phosphorylation of tyrosine 41 of H3 (by JAK2 tyrosine kinase) has been shown at active promoters where it correlates with H3K4me3 as well as over the transcribed regions of tissue-specific active hematopoetic genes (Dawson et al., 2012).

1.2.3.4. Other histone modifications

Mono-ubiquitylation of H2B has been correlated with both transcriptional activation and repression of genes (Wright et al., 2012). Ubiquitylation of H2A (H2Aub) is involved in polycomb-mediated repression (Endoh et al., 2012). The RING1B component of the PRC1 complex can mediate chromatin compaction and gene silencing independently of histone ubiquitylation activity (Endoh et al., 2012; Eskeland et al., 2010a). Other modifications not discussed here include biotinylation, citrullination, carbonylation, poly(ADP-ribosyl)ation, and sumoylation, amongst others.

1.2.4. Chromatin Organization in the nucleus

Rather than being randomly organized chromatin in the nucleus is spatially organised and is thought to affect gene function. Nuclear components have specific localisations within the nucleus and movement of chromatin into environments that are conducive for transcription and post-transcriptional processing to take place is key for normal gene function.

1.2.4.1. Chromosome Territories

Chromosomes occupy specific positions within the nucleus known as chromosome territories (Cremer et al., 2006; Heard and Bickmore, 2007). Gene-rich chromosomes and chromatin regions are generally located towards the interior and gene-poor chromosomes towards the nuclear periphery (Figure 1.14 and section 1.2.2.4) (S Boyle et al., 2001; Croft et al., 1999). Chromosome territories are not completely spatially separated as

intermingling of chromosomes occurs at the territory periphery (Branco and Pombo, 2006). Furthermore, within these territories gene rich and poor regions can be spatially separated (Shopland et al., 2006; Simonis et al., 2006). Translocations that join together chromosomes that normally are found in different nuclear environments result in altered nuclear organisation and gene expression (Harewood et al., 2010). There are some reports that this can alter cancer risk and the full impact of this area of nuclear organisation on tumours with complex karyotypic rearrangements is yet to be assessed.

The human major histocompatibility complex (MHC) was the first example of genes looping out of their chromosome territory when active (Volpi et al., 2000). The interferon-γ induced looping out of MHC genes is mediated through the transcriptional activator P-STAT1 and the chromatin remodeller BRG1 (Christova et al., 2007). The *Hox* genes have also been shown to loop out of their chromosome territories when active during differentiation (Chambeyron and Bickmore, 2004). Therefore it is possible to increase the surface of chromosome territories by looping out into the intrachromosomal compartment or conversely by infoldings into these territories (Cremer et al., 2006; Heard and Bickmore, 2007). The scale of looping out of gene-rich genomic regions from chromosome territories has been recently revealed by FISH (Boyle et al., 2011). Regions looped out from chromosome territories have an enhanced probability of interacting with sequences from other chromosomes - as measured by chromosome conformation capture (3C) methods (Kalhor et al., 2012; Würtele and Chartrand, 2006). It is thought that the relocalisation of genomic regions in their chromosome territories allows access to different nuclear environments containing transcriptional machinery components (Brown et al., 2006; Moen et al., 2004; Morey et al., 2009; Osborne et al., 2004).

1.2.4.2. Chromatin folding

Fluorescence in situ hybridisation (FISH) is a molecular cytogenetic technique which has been used by the Bickmore lab (and others) to detect whole chromosomes as well specific DNA loci in the nucleus. FISH can be used to measure the compaction of chromatin over a specified genomic region. A random walk model of the behaviour of higher order chromatin has been identified using FISH data with pairs of probes at different genomic

intervals along the linear DNA molecule (van den Engh et al., 1992). With probe separations of 100kb to 1.5 Mb there is a linear relationship between the mean inter-probe distances (squared) and the genomic distance that fits a "random walk" model with the chromatin behaving like a flexible polymer (van den Engh et al., 1992). This model was also suggested in subsequent work using probes for 150kb to 190Mb intervals on chromosomes 4, 5 and 19 and modelling of the large-scale chromatin geometry (Sachs et al., 1995; Yokota et al., 1995). This analysis produced a "random walk/giant loop" model of large-scale chromatin organisation whereby at megabase intervals the chromatin is organised in giant loop structures which are fixed (or "tethered") on a flexible backbone that shows random walk behaviour (Sachs et al., 1995). In this model interprobe distance conform to a Rayleigh distribution and are characterised by a standard deviation/mean ratio of ~0.52 and a median/mean ratio of ~0.94.

Chromatin compaction using pairs of FISH probes has been well studied using this method during differentiation of embryonic stem cells, (Chambeyron and Bickmore, 2004), in the developing embryo (Chambeyron et al., 2005; Morey et al., 2007), comparing wildtype and mutant cells (Ragnhild Eskeland et al., 2010) and also for the study of multiple loci in the same cells (Gilbert et al., 2004). Similar data distributions are observed for multiple regions of the mouse and human genomes in a wide variety of cell types and in tissue sections (Chambeyron and Bickmore, 2004; Gilbert et al., 2004; Williamson et al., 2012). Furthermore data from 2D FISH analysis of fixed cells has been shown to recapitulate the chromatin compaction observed with 3D FISH though distances are typically larger (Ragnhild Eskeland et al., 2010; Morey et al., 2007).

The degree to which chromatin in compacted is not constant over the genome. Regions with low gene densities (as in G-bands) are significantly highly compacted compared with high gene density regions (as in R-bands) (Gilbert et al., 2004). However, "open" and "closed" chromatin status does not absolutely distinguish between regions that are actively transcribed or being repressed - as active genes can be found in compact regions and silent genes in open regions (Gilbert et al., 2004). It is thought that the decompact "open" chromatin structure facilitates active transcription as it is permissive to access by transcriptional machinery but does not necessitate that genes will be expressed (Gilbert et al., 2004).

1.2.4.3. Transcription factories

Active genes co-localise at sites of transcription measuring approximately 80nm in diameter which are known as "transcription factories" (Osborne et al., 2004; Sutherland and Bickmore, 2009). These foci can be visualised by incorporation of bromouridine (BrdU) into newly synthesized RNA (Jackson et al., 1993). Between 4 and 20 RNA polymerase II molecules have been shown to cluster together in these transcription factories at a time (Iborra et al., 1996). Co-regulated genes have been shown to cluster in space, beyond their observed linear organisation by gene order in the genome and that this spatial co-localisation of genes leads to higher expression (Ben-Elazar et al., 2013; Rieder et al., 2012; Sutherland and Bickmore, 2009). The co-localisation of active genes at the transcription factories has been suggested to be important in the generation of specific cancer-associated translocations (Osborne et al., 2007).

Genes within their chromosome territory have been shown to relocate to the edges and outside of their territories to associate with transcription factories (Morey et al., 2009). Furthermore nuclear hormone receptors are found to concentrate at transcription factories when hormone ligand is bound (Sutherland and Bickmore, 2009). Recent evidence has also suggested that transcription factories are the result of the spontaneous concentration of active chromatin modifications like histone H4 acetylated at Lysine 16 (H4K16ac) which induce the formation of stiff microdomains (Canals-Hamann et al., 2013).

1.2.4.4. The Nuclear Periphery

Gene density is a major factor in the localisation of chromosomes in nuclear space. Chromosomes with a high gene density ("gene-rich" chromosomes) occupy a central position in the nucleus and conversely those that have a low gene-density ("gene-poor" chromosomes) are located at the periphery (S Boyle et al., 2001; Croft et al., 1999) (Figure 1.13). This radial distribution of gene-rich and gene-poor chromosomes is also present in tumour nuclei though is it less pronounced compared to normal nuclei (Cremer et al., 2003) which could be due to the presence of chromosomal rearrangements such as translocations in tumour cells (Harewood et al., 2010). Proximity to the nuclear periphery is also associated with the level of transcription, with repressed genes located closer to the periphery than more centrally located active gene clusters. Moreover,

tethering chromosomes to the nuclear periphery using a *E.coli* lac operator system (lacO) can reduce the expression of some endogenous genes located near the lacO sites (Finlan et al., 2008). A notable exception to this dogma of nuclear periphery as a repressive-compartment is in mouse retinal rod cells where this is inverted: gene-poor heterochromatin is packed in the centre of the nucleus and gene-rich euchromatin at the periphery (Solovei et al., 2009). However, overall chromatin at the nuclear periphery represents inactive, compact and gene-poor chromatin.

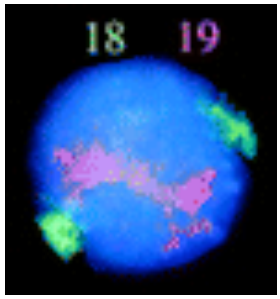


Figure 1.13: Chromosome territories in interphase nuclei for human chromosomes 18 and 19. Human lymphoblastoid cell nucleus hybridized by FISH with paints for the gene-rich chromosome 19 (red) and gene-poor chromosome 18 (green) showing the radial organization of these chromosomes in the nucleus. (Figure taken from Croft et al., 1999)

Tethering of the *E. coli* DNA adenine methyltransferase (Dam) to a chromatin protein, allows *in vivo* targeting of Dam to binding sites of the protein, which become marked by DNA A methylation (Steensel and Henikoff, 2000). Dam ID technology has been used to target Dam to proteins in the nuclear lamina providing a map of chromatin domains that associate with the nuclear periphery. These lamina-associated domains (LADs) span mega-base sized regions in humans and mice, where they generally correlate with gene-poor regions that are transcriptionally inactive and late replicating (Guelen et al., 2008; Peric-Hupkes et al., 2010). A role for the nuclear lamina in control of gene expression is further demonstrated by dissociation of genes from the lamina which results in either their concordant activation or keeps them poised for transcription (Peric-Hupkes et al., 2010).

1.2.4.5. Chromatin interactions and topology

There is increasing evidence of elements of the genome physically interacting over large genomic distances with one another, indicating that a three-dimensional network of genomic organisation exists which is defined by these physical interactions (Dostie and Dekker, 2007; Dostie et al., 2006). Development of techniques such as Chromosome Conformation Capture (3C), which cross-link together chromatin regions in proximity in the genome has allowed a number of long range *cis* and *trans* associations to be identified. For 3C chromatin interactions are fixed using formaldehyde before being digested with a restriction enzyme (Dekker et al., 2002). The digested fragments retain the cross-links created by formaldehyde and the sticky ends are ligated before the cross-links are reversed (Dekker et al., 2002). The ligation products are detected by qPCR (quantitative polymerase chain reaction) amplification of selected ligation junctions using primers for the genes of interest designed towards the end of the restriction fragments (Dekker et al., 2002). Since the advent of 3C a number of techniques have also evolved based on this method for identifying chromatin associations (Figure 1.14).

Using the 3C technique a number of *cis* interactions connecting promoters and regulatory regions have been identified including at both the mouse and human β-globin locus (Deng et al., 2012; Tolhuis et al., 2011), the α-globin locus (Vernimmen et al., 2007) and the TH2 cytokine locus (Spilianakis et al., 2005). The importance of the enhancer-promoter chromatin loop has been demonstrated by forcing the formation of a loop between the locus control region (LCR) and promoter of the β-globin locus, which resulted in recruitment of RNA polymerase II and activation of transcription (Deng et al., 2012). The effect of enhancers of gene activity can be attenuated by insulator sequences that are often bound by CCCTC-binding factor (CTCF) (Phillips and Corces, 2009).

CTCF is a ubiquitously expressed, highly conserved, zinc-finger DNA-binding protein with many roles including transcriptional activation, transcriptional repression, as a transcription factor involved in hormone-responsive gene silencing, as an insulator protein, as a protein involved in imprinting/X-chromosome inactivation, as a boundary element and as an aid to long-range chromatin interaction within and between chromosomes (Phillips and Corces, 2009). CTCF has been shown to bind sites around the

mouse β-globin locus that spatially cluster in the erythroid cell nucleus and mediate long-range DNA looping at this locus with maintenance of local histone modifications (Splinter et al., 2006). It is well reported that cohesin, a protein complex that holds sister chromatids together, localises to CTCF binding sites and that association of this complex is essential for the insulator properties of CTCF (Parelho et al., 2008; Rubio et al., 2008; Stedman et al., 2008; Wendt et al., 2008). However, cohesin and the transcriptional-coactivator mediator can also co-occupy enhancer and active promoter sites independently of CTCF where they are thought to maintain DNA loops by the formation of ring structures connecting the two sites (Kagey et al., 2010)

Circular 3C (4C) is a development of the 3C method, which is used to identify interactions between a region of interest with any other genomic region. As shown in Figure 1.14 the 3C ligated fragment is digested, then re-ligated to form circularised fragments and any interacting regions can be identified by hybridisation to microarrays (Simonis et al., 2006) or next generation sequencing (Splinter et al., 2011). 4C interaction profiles of genes located in inactive (β-globin) and active (Rad23a) chromatin domains are very different from one another with many contacts identified for Rad23a with active regions but expression-dependent active/inactive contacts for β-globin in different tissues (Simonis et al., 2006). 4C has also been used to identify contacts in response to glucocorticoid receptor (GR) stimulation where no large-scale movement was observed (Hakim et al., 2011) and *cis* contacts between genes repressed in association with polycomb group proteins (Bantignies et al., 2011; Tolhuis et al., 2011).

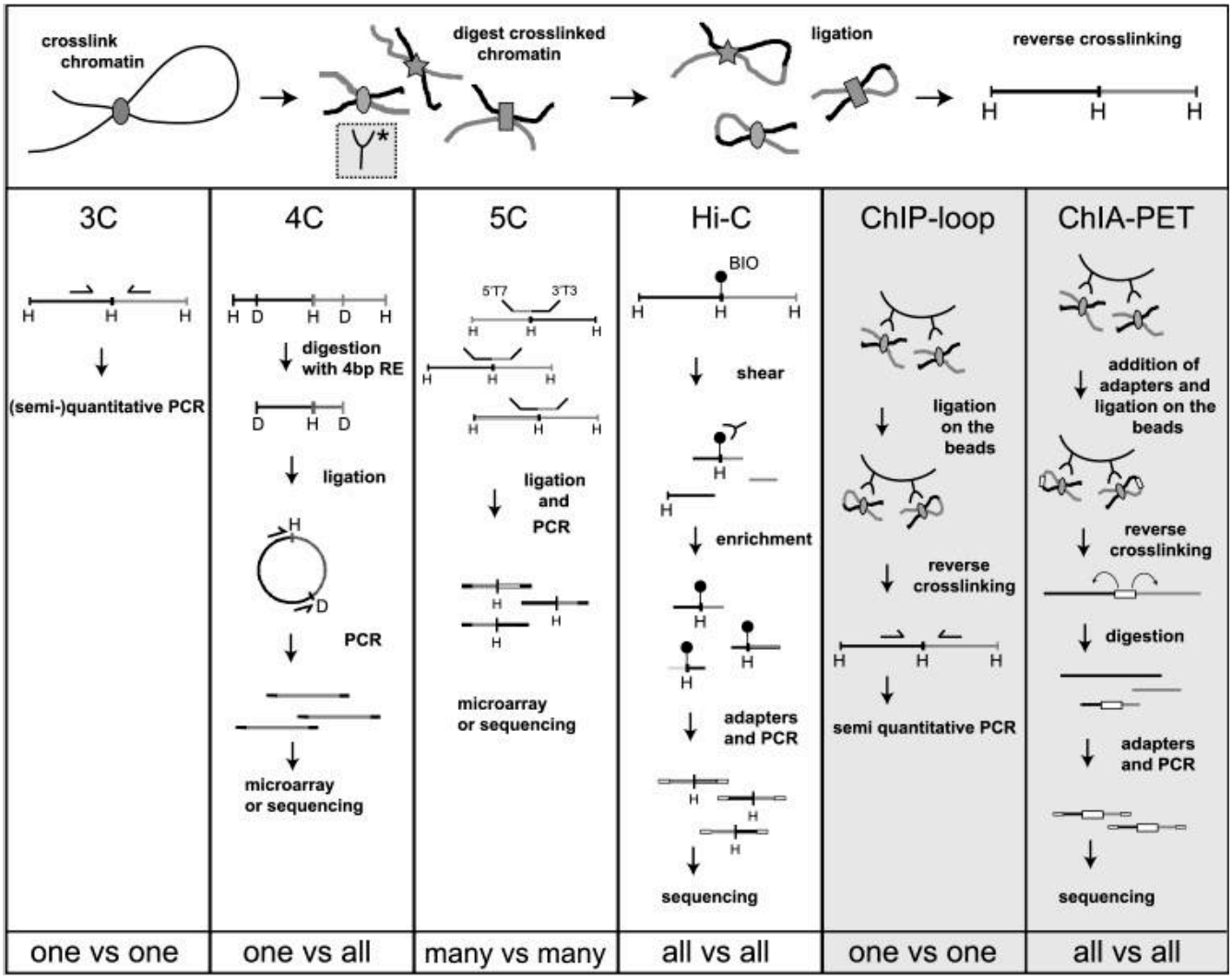


Figure 1.14: Schematic overview of 3C-derived methods. The horizontal panel of the "C" methods. The vertical panels indicate the steps that are specific to separate methods. Figure taken from (de Wit and de Laat, 2012)

5C (3C carbon copy) is a high-throughput adaptation of 3C used to detect long-range interactions between multiple genomic loci. 5C uses pools of reverse (T3 tagged) and forward (T7 tagged) primers that are designed to detect all possible junctions in a 3C library of cross-linked restriction fragments, which are then ligated to form a carbon-copy library which is amplified with primers (to T7 and T3) and sequenced or hybridised to microarrays (Dostie et al., 2006). Analysis of the HOXA gene cluster showed long-range interactions between the active components of the gene cluster in cell lines (Wang et al., 2011) and the formation of active/inactive contact domains along the developing body axis *in vivo* (Noordermeer et al., 2011). This does not however marry up with the

observations by FISH that active HOX loci are in a decondensed chromatin conformation and therefore not clustered together (Dostie and Bickmore, 2012).

5C and super-resolution microscopy have been used to analyse the spatial organisation of a 4.5 Mb domain (containing *Xist*) and led to the discovery of 200 kb - 1 Mb topologically associating domains (TADs) that correlate with H3K27me3 or H3K9me2 blocks, lamina-associated domains and coordinately regulated gene clusters (Nora et al., 2012). Furthermore the TADs were able to spatially segregate active/inactive chromatin regions and disruption of the TAD boundaries led to long-range misregulation of transcription (Nora et al., 2012)

Hi-C (Lieberman-Aiden et al., 2009) can identify long-range interactions which occur across the genome. After formaldehyde crosslinking of interacting DNA and restriction digestion the fragment ends are biotin-labelled before ligation creating a library of products with biotin at the junction. This library is sheared and the biotin labelled junctions pulled down by streptavidin beads and subjected to high-throughput sequencing to identify the interacting regions. Similarly to Nora et al (2012), partitioning of the genome into large topological domains (~1 Mb) that are highly conserved and stable across cell types has been identified using Hi-C detection of chromatin interactions (Dixon et al., 2012). Furthermore the boundaries of the TADs identified by Dixon and colleagues were enriched with CTCF , transcription start sites (TSS), transfer RNAs, housekeeping genes and short interspersed element (SINE) retrotransposons, which are thought to be involved in TAD structure (Dixon et al., 2012).

The ChIP-loop assay combines 3C and ChIP to identify genomic interactions that are associated with a particular protein (Horike et al., 2005). This technique uses the purified cross-linked chromatin after digestion with restriction enzymes (3C library) for precipitation with protein A/G beads bound with antibodies for the protein of interest. This fraction of protein-bound cross-linked fragments are ligated and analyzed by qPCR. ChIP-loop was used to study interactions of a methyl-binding protein (MECP2) which mediates a silent chromatin structure (11kb DNA loop) that is lost in Rett syndrome where MECP2 is mutated (Horike et al., 2005). Combined 3C-ChIP-cloning (6C) uses a

3C library to identify fragments enriched in a protein of interest using a specific antibody (as with ChIP-loop) by reversing cross-links in enriched fragments cloning them into a vector which can be screened by digestion and sequencing to identify interaction partners (Tiwari et al., 2008). This has been used to identify a number of long-range inter/intra-chromosomal contacts mediated by the polycomb group protein EZH2, which correlated with transcriptional repression (Tiwari et al., 2008).

Chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) also allows the detection of chromatin interactions associated with a protein of interest and allowed a genome-wide interactome to be identified for the ER (Fullwood et al., 2009) and CTCF (Handoko et al., 2011). For ChIA-PET cross-linked DNA fragments are enriched by ChIP and tethered by covalent linkers before extraction (by restriction digestion/streptavidin-magnetic beads) of paired-end tags (PETs) for sequencing (Fullwood et al., 2009; Goh et al., 2012). The interactome map identified for the ER in breast cancer cells (MCF7s) suggested a mechanism for coordinated expression of genes by the formation of intrachromosome loops (consisting of "anchor" and "loop" genes) that are maintained by the ER (Fullwood et al., 2009). Similarly the CTCF interactome in mouse embryonic stem cells (ESCs) showed a network of cross-linking between regulatory elements and promoters anchored by CTCF (Handoko et al., 2011).

1.2.4.6. The Estrogen Receptor

Estrogen action is mediated via the estrogen receptors alpha (ERα) and beta (ERβ), which are part of the nuclear receptor superfamily. ERα and ERβ are both expressed in normal mammary epithelial cells but can also be differentially expressed in different cell types, for example ERβ is expressed in the prostate epithelium and ERα in stroma (Heldring et al., 2007; Nilsson and Gustafsson, 2002) where it promotes carcinogenesis (Ricke et al., 2008). Knockout studies in mice showed that ERα is crucial for formation of ducts during mammary development (Hewitt et al., 2005) whereas ERβ knockout leads to much milder mammary phenotypes and an increase in Ki-67 expression in the luminal epithelial cells of the mammary gland (Förster et al., 2002). ERβ is also thought to have a protective role and is often silenced in malignant breast and prostate cancer (Järvinen et al., 2000; Zhu et al., 2004).

The ERs regulate gene expression via a sequence of events initiated by ligand binding to ERα or ERβ to induce dimerization as either homodimers (ERα/ERα or ERβ/ERβ) or heterodimers (ERα/ERβ). These then translocate to the nucleus where they bind to specific estrogen response elements (EREs) in the promoters of target genes and trigger the recruitment of cofactors and RNA polymerase II. ERα and ERβ have different biological functions, the ERα homodimer promotes cell growth and proliferation whereas ERβ activation is anti-proliferative and promotes apoptosis (Helguero et al., 2005). Forced expression of ERα/ERβ heterodimers in breast and prostate cancer cells leads to growth inhibition (Powell et al., 2012) however evidence for the existence of such heterodimers *in vivo* is lacking. The anti-tumourigenic protective effect of ERβ is also demonstrated in ER+ve breast tumours where ERβ expressing patients show better prognosis and rate of disease-free survival (Omoto et al., 2001). ERα and ERβ show distinct binding profiles in the genome and are thought to regulate different genes (Leitman et al., 2010).

Transcriptional activity of genes through recruitment of coregulator complexes has mainly been studied with ERα and may be different in ERβ mediated transcription. Ligand binding to ER leads to conformational changes in the receptor, binding to estrogen response elements (EREs) at promoters and extensive recruitment of coregulatory proteins involved in transcription and chromatin remodelling. The coregulatory proteins recruited to ERα and transcriptional activity of target genes are determined by the ligand bound to the receptor (Figure 1.15 (Teyssier et al., 2010)). Coactivators are able to modify chromatin and facilitate transcription upon estrogen stimulation. The SRC (p160) family of coactivators are recruited in response to estrogen binding to ERα and include SRC1 (NCOA1), SRC2 (NCOA2/TIF2) and SRC3 (NCOA3/AIB1) (Heldring et al., 2007; Zwart et al., 2011). The SRC coactivators also recruit histone methyl transferases (PRMT1 and CARM1/PRMT4) to ERα (Teyssier et al., 2010). PRMT1 methylates arginine 3 on histone H4 (H4R3) and facilitates deacetylation of H4 tails by p300 and active transcription (Wang et al., 2001). The coactivators CBP and p300 are histone acetyltransferases (HATs) which interact with SRC family members to enhance transcription mediated by estrogen bound ERα (Green and Carroll, 2007). Furthermore the expression of SRC3 and CBP bound genes can be used for prediction of patient response to endocrine therapy (Zwart et al., 2011) The

ATP-dependent remodelling complex SWI/SNF also acts as a coactivator through spatial reorganisation of nucleosomes which allows the Mediator complex and transcriptional machinery to be recruited to the initiation start point (Kassabov et al., 2003; Teyssier et al., 2010).

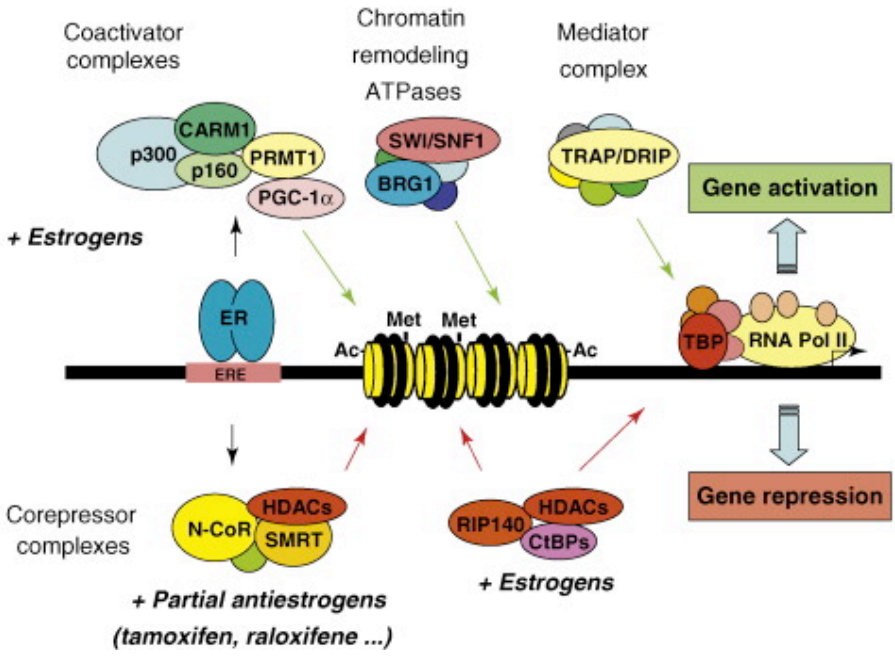


Figure 1.15: The ERα Coregulator Complexes. The recruitment of corepressor or coactivator complexes depends on the type of ligand bound to ERα. (taken from (Teyssier et al., 2010))

By contrast, upon antiestrogen ligand binding corepressor complexes are recruited that modify chromatin making it less accessible transcriptional machinery and block transcription. NCOR1 and SMRT (NCOR2) are corepressor complexes which are involved in the recruitment of histone deacetylases (HDACs) which are key to modifying the chromatin so that access to the promoters is inhibited (Dobrzycka et al., 2003; Teyssier et al., 2010). In response to estrogen the ER can bind with RIP140 which interacts with HDACs and acts as a corepressor (Teyssier et al., 2010).

As discussed in section 1.2.4.5. identification of the ERα interactome in estrogen-responsive cells showed that a number of complex intrachromosomal interactions exist that are mediated by estrogen for active transcription (Fullwood et al., 2009). Further evidence for large-scale changes in chromatin architecture comes from tethering the ER to DNA resulting in large scale decondensation of chromatin (Nye et al., 2002; Sharp et

al., 2006). It has also been reported that estrogen stimulation of breast cancer cells leads to the recruitment of PAD2 enzyme which targets ERα bound promoters and citrullinates H3R26 leading to chromatin decondensation and activation of transcription (Zhang et al., 2012).

Through genome-wide binding studies it is apparent that the ER can regulate gene expression from a distance (Carroll et al., 2006; Hurtado et al., 2011). FoxA1 binding is highly enriched at distal gene promoters and it is hypothesized that interactions between distal binding sites and elements like FoxA1 could potentiate DNA looping (Carroll et al., 2005). FoxA1 is a pioneer factor that binds to compact chromatin and creates an open chromatin state by displacing linker histone H1 and disrupting the H3/H4 tetramer (Cirillo et al., 2002, 1998). Moreover FoxA1 binding has been found to correlate with regions rich in H3K4me1/2 and poor in H3K9me2 (Lupien et al., 2008). A number of other pioneer factors have been associated with hormone-dependent cancer including PBX1, TLE and GATA3 (Jozwik and Carroll, 2012).

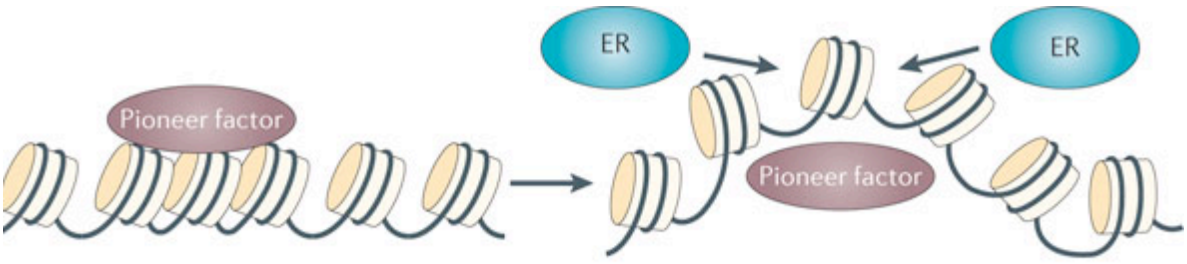


Figure 1.16: The function of pioneer factors. By binding with condensed chromatin pioneer factors are able to increase accessibility to estrogen response elements (EREs) facilitating ERα-chromatin interactions (Jozwik and Carroll, 2012).

Almost all ER-chromatin interactions which effect gene expression are dependent on FoxA1 and are negatively regulated by CTCF (Hurtado et al., 2011). Comparison of CTCF and ER binding data has shown that CTCF binding can divide the genome into blocks of estrogen-regulated genes with ER binding sites (Chan and Song, 2008). Moreover CTCF binding upstream of FoxA1 has been shown to demarcate the estrogen regulated genes by acting as boundary to the spread of heterochromatin (Zhang et al., 2010).

Estrogen or tamoxifen treatment has no effect of the genome-wide CTCF binding events in breast cancer cell lines (Ross-Innes et al., 2011). In fact CTCF was co-bound with key transcription factors some of these CTCF binding events are cell-line specific and are associated with highly expressed genes. However numerous binding sites exist that share ER- and cohesin-binding events but lack CTCF and are regulated by estrogen (Schmidt et al., 2010).

In the absence of estrogen, breast cancer cells are growth arrested and can only proliferate and re-enter the cell cycle after estrogen stimulation (Prall et al., 1997), and this is cohesin-dependent (Schmidt et al., 2010). The fact that cohesin has a role to play in ER-mediated gene expression was also highlighted by a genome-wide functional screen in which silencing of NIBL, RAD21, SMC3 (components of the cohesin complex) conferred tamoxifen resistence in ER+ve breast cancer cells (Mendes-Pereira et al., 2012). Another key regulator of the cell cycle is FoxM1 which was recently shown to co-bind with the ER (like FoxA1) in breast cancer cells where it interacts with co-activator associated arginine methyltransferase (CARM1) and is involved in transcriptional activity of ERα regulated genes (Sanders et al., 2013).

1.3 Cancer and Epigenetics

The term “epigenetics” is used loosely in the literature with a various different meanings (as discussed in (Bird, 2007)). The term was first used by Conrad Waddington to describe the relationship between genotype and phenotype during development (Waddington, n.d.). Later definitions that are commonly quoted include those by Robin Holliday and Arthur Riggs, where epigenetics is used to describe heritable changes that cannot be explained by the DNA sequence itself (Holliday, 1990; Riggs, 1996). The current use of the term “epigenetics” however deviates from these historical definitions in that it is used to describe changes that are not necessarily heritable but give rise to “altered activity states” (Bird, 2007). At a Cold Spring Harbour meeting in 2008 a consensus definition for an epigenetic trait was defined as a "stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence” (Berger et al., 2009) . In this thesis I have used the term “epigenetics” to describe changes in gene expression that are not associated with changes to the DNA sequence but to changes in chromatin structure, organization and modifications to the DNA that do not affect the sequence. Here I talk about the role of epigenetics in tumourigenesis with particular regard to breast cancer.

1.3.1. DNA methylation in breast cancer

Aberrant patterns of DNA methylation have been reported in cancer cells with both gains (hypermethylation) and losses (hypomethylation) of methylation. Global DNA hypomethylation is a consistent characteristic of tumours (Bernardino et al., 1997) and occurs at repetitive elements and at satellite DNAs, which are heavily methylated in normal cells. Many hypomethylated genes are normally expressed exclusively in germline cells and have been denoted as “cancer-germline” (CG) genes i.e. genes that rely primarily on DNA methylation for their tissue-specific expression (De Smet and Loriot, 2013). Discrete blocks of hypomethylation identified across the genome in tumours associate with highly variable expression profiles that are thought to contribute to tumour heterogeneity (Hansen et al., 2011). They also correspond to domains which show patterns of partial methylation in normal tissues (Hon et al., 2011; Berman et al.,

2011) A number of cancer-specific differentially methylated regions (DMRs) - highly variable between tumour types – have been attributed to the loss of defined boundaries to DNA methylation around CpG islands (Hansen et al., 2011)..

As well as large domains of hypomethylation, focal DNA hypermethylation has been reported at CGIs in tumours. For example, this has been implicated in the loss of expression of critical genes ultimately leading to the progression of breast tumours and includes the genes encoding the estrogen and progesterone receptors in approximately half of breast cancer cases (Lapidus et al., 1996; Ottaviano et al., 1994). Hypermethylation of the genes involved in cell cycle regulation, *p16* and *14-3-3-sigma*, has also been demonstrated in breast cancer (Ferguson et al., 2000; Herman et al., 1995). In cases of sporadic breast cancer, where *BRCA1* mutations are not observed, loss of *BRCA1* expression has been correlated to DNA methylation at the promoter of this gene (Bal et al., 2012; Dobrovic and Simpfendorfer, 1997). Inactivation of *BRCA1* associated with promoter hypermethylation has also been seen in ovarian cancer and lung adenocarcinoma (Botana-Rial et al., 2012; Wang et al., 2012). Other genes frequently hypermethylated in breast can include Glutathione S-transferase (*GSTP1*) which protects against cytotoxic and carcinogenic agents (Esteller et al., 1998); E-cadherin which suppresses tumour invasion and metastasis (Graff et al., 1995); and Tissue inhibitor of metalloproteinase-3 (*TIMP-3*) which antagonizes matrix metalloproteinase activity and can suppress tumour growth, angiogenesis, invasion, and metastasis (Bachman et al., 1999).

The assumption implicit in these studies is that the DNA hypermethylation is directly causative of the gene silencing and that it is a driver of tumourigenesis through the silencing of tumour suppressor genes. However a careful analysis of tumour ontology and analysis of the 'normal' cells from which tumours arise has suggested that this idea needs to be revised. In breast (Sproul et al., 2011) and other cancers (Sproul et al., 2012) hypermethylated genes were found to be already silent in the normal tissues from which the tumours arose, so that DNA methylation was acquired subsequent to gene repression rather than being a cause of it. The acquisition of DNA methylation as a consequence, rather than a cause, of gene silencing is also evident from the recent ENCODE data release where transcriptor factor binding is thought to block the acquisition of DNA methylation (Ecker et al., 2012).

1.3.2. Histone Acetylation and Methylation in breast cancer

In breast cancer, 'aberrant' histone modifications have also frequently been associated with silencing of tumour suppressor genes and genomic instability. The regions of partial (hypo)methylation discussed above show a correlation with both H3K9me3 and H3K27me3 and, interestingly, DNA methylation and these repressive histone modifications were allele specific and mutually exclusive to form domains of gene silencing (Hon et al., 2012). Similarly, it has also been shown that genes subject to DNA hypermethylation in tumours tend to be repressed by the polycomb system (H3K27me3) in the normal cells from which the tumours arose (Sproul et al., 2012).

H3K27me3 is associated with increased risk of cancer in colon cancer models as well as stem and progentitor cells (Ohm et al., 2007; Schlesinger et al., 2007). Overexpression of EZH2 (the H3K27 HMTase) occurs in a number of malignancies, especially in those that are steroid dependent like breast cancer (Kleer et al., 2003). In normal breast, levels of EZH2 and H3K27me3 in luminal cells change dramatically during pregnancy, and Ezh2 is required for alveologenesis, and progenitor cell proliferation (Pal et al., 2013). In human mammary epithelial cells, overexpression of EZH2 leads to an increase in anchorage-independent growth and cell invasion (Bracken et al., 2003; Kleer et al., 2003) and in basal-like tumours EZH2 promotes a gene expression programme in which both basal- and luminal-lineage genes are co-expressed (Granit et al., 2012). EZH2 is also involved in regulating estrogen-dependent transcription through a novel binding partner REA (repressor of estrogen receptor activity) which acts as a corepressor of the ER (Hwang et al., 2008).

HDAC inhibitors can alter chromatin structure to allow re-expression of key genes that have been silenced in tumours in association with apoptosis/growth inhibition of the tumour cells (Khan and La Thangue, 2008; Marson, 2009). In ER –ve breast tumours, the HDAC inhibitors Scriptaid and Trichostatin A (TSA) promote increased of H3 and H4 acetylation and re-expression of ER, leading to growth inhibition (Keen et al., 2003). Combination treatments with HDAC and DNA methyltransferase (DNMT) inhibitors have shown better clinical responses than either type of drugs alone. HDAC/DNMT inhibitor combinations have shown better re-expression of aberrantly silenced genes in a

number of cancer cell lines including breast as well as increased apoptosis of tumour cells (Belinsky et al., 2003; Cameron et al., 1999).

A model of the role of histone methylation and the enzymes involved is shown in Figure 1.17 (From review (Huang et al., 2011)).

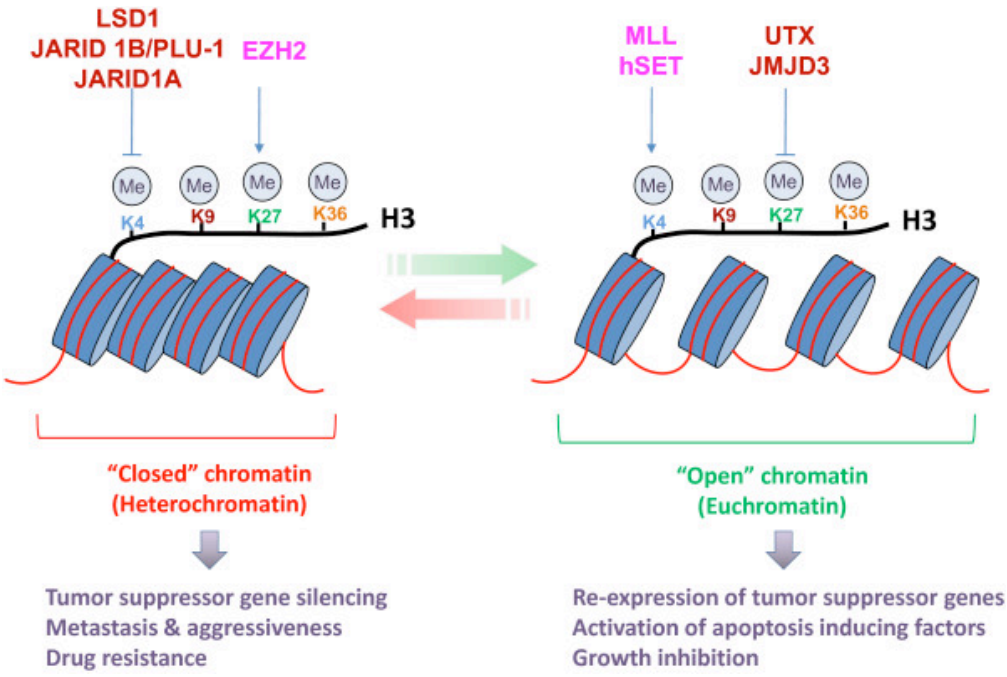


Figure 1.17: Model of dynamic interplay of enzymes mediating methylation of histone lysines. Methylases are shown in pink and demethylases are shown in brown. (From Huang et al 2011 review).

The long non-coding RNA (lincRNA) *HOTAIR* is highly expressed in primary tumours and metastases where it has been linked to PRC2 recruitment and subsequent deposition of the H3K27me3 repressive mark **(Chisholm et al., 2012; Gupta et al., 2010)**. Moreover it is possible to inhibit cancer invasiveness by depletion of *HOTAIR* and this effect was most prominent in cells with high PRC2 activity **(Gupta et al., 2010)**. *HOTAIR* transcription is estrogen-mediated in breast cancer where it is thought to contribute to tumour progression **(Bhan et al., 2013)**.

Histone lysine-specific demethylase 1 (LSD1) is involved in demethylating lysine 4 of histone H3 (H3K4me1 and H3K4me2) in association with transcriptional activation (Shi et al., 2004). LSD1 is highly overexpressed in ER-ve breast tumours and its inhibition prevents tumour cell growth, making it a therapeutic target for this aggressive subtype of

breast cancer (Lim et al., 2010). There is also mechanistic evidence for gene activation by interplay between ligand-bound nuclear receptors and histone methylation. In the ER-positive breast cancer cell line (MCF7) genes with promoters co-bound by both LSD1 and ERα could be stimulated by treatment with estrogen (Garcia-Bassets et al., 2007).

LSD1 has also been demonstrated to have a complex relationship with HDAC inhibitors in human breast cancer cells. HDAC treatment actually causes increased H3K4me2 (the substrate of LSD1) whereas inhibition of LSD1 with Pargyline or siRNA causes increased acetylation of H3K9 (Huang et al., 2012). When combined these HDAC/LSD1 inhibitor treatments cause increased H3K4me2 and H3K9Ac with growth inhibition of breast tumour cells, indicating a cooperative strategy for regulating gene expression (Huang et al., 2012).

1.3.3. The nuclear periphery and cancer

The nuclear periphery (see section 1.1.2.4.4). is generally associated with transcriptional silencing. Importantly the domains of DNA hypomethylation identified in colorectal tumors correlate with late replication and nuclear lamina attachment (LADs) (Berman et al., 2012). This has also been seen in breast cancer (Hon et al., 2012) (Figure 1.18) This suggests that the genome-wide changes in DNA methylation observed in cancer are linked to silencing mechanisms regulated by chromatin organization and the nuclear periphery, but the direct involvement of the nuclear periphery in tumourigenesis has not been investigated directly.

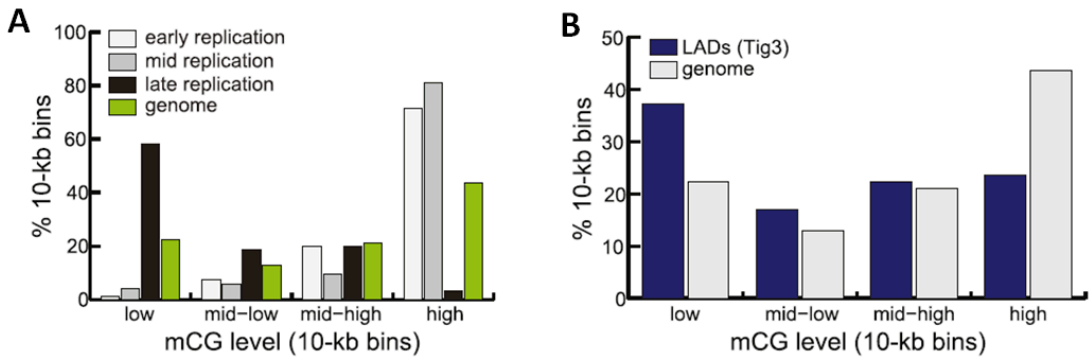


Figure 1.18: Domains of hypomethylation correlate with late replication and the nuclear lamina in breast cancer. (A) Distribution of HCC1954 %mCG for 10-kb regions that are consistently early-, middle-, and late-replicating in four cell types, compared to the background genome (green). (B) Distribution of HCC1954 %mCG for 10-kb regions that are found in Tig3

lamina-associated domains (blue), compared to the background genome (gray). (Figure taken from Hon et al., 2012)

1.3.4. Long-range epigenetic silencing (LRES) in cancer

Misregulation of gene expression correlating with genetic and epigenetic lesions in tumourigenesis has been largely explored at the single gene level. However, there are increasing reports of contiguous genes being co-ordinately repressed in association with tumour progression - a phenomena which has become known as Long Range Epigenetic Silencing (LRES). One of the first examples of this was a 100kb region containing the HoxA gene cluster which was silenced in association with loss of active histone modifications and aberrant DNA methylation in breast cancer (Novak et al., 2006).

In colorectal cancer, concordant gene silencing of a 4Mb region on chromosome 2q14.2, accompanied by DNA hypermethylation of a subset of CGIs in that region, was associated with repression of genes within the entire chromosome band despite CGIs at the promoters of many of these neighbouring genes remaining unmethylated (Frigola et al., 2006). LRES at this locus was also associated with the repressive histone modification H3K9me2 (Frigola et al., 2006). Recently this same genomic region has also been shown to be repressed by LRES in prostate cancer in association with hypoacetylation of H3K9, H3K9me2, H3K27me3 and localised DNA hypermethylation (Devaney et al., 2011).

Since this first seminal example of concordant silencing of a large chromosomal region in association with DNA hypermethylation in cancer, other genomic regions or "agglomerates" of hypermethylated regions have been described. Analysis of DNA methylation in breast cancer specimens and normal breast tissue identified large chromosomal regions (up to ~700 kb) of cancer-specific LRES which occurred frequently at gene clusters including the protocadherin family cluster (PCDH) on chromosome 5q31 (Novak et al., 2008). This region has also been identified as an LRES hypermethylated domain in colorectal adenomas and carcinomas where methylation of specific genes in the region mark progression along the adenoma-carcinoma transition (Dallosso et al., 2009). This region was also found to be hypermethylated in Wilms'

tumour (Dallosso et al., 2012). Additionally, LRES associated with DNA hypermethylation at a 2.7 Mb region on chromosome 12q14 has been identified in gastric cancer (Park et al., 2011). Cancer-associated hypermethylation of *Ikaros*, which belongs to a family of genes encoding transcription factors involved in regulating cell differentiation, has also been shown to be embedded within a region described as LRES due to association with DNA hypermethylation and loss of H3K4me3 (Javierre et al., 2011).

Computational biology has been successfully employed to systematically identify LRES through the integration of trancriptomic and epigenomic data for different types of tumours. In bladder carcinoma a sliding window approach was used to generate scores for how well as gene's expression correlated with that of its neighbours (Stransky et al., 2006). This allowed regions of co-ordinately down-regulated genes to be identified and regions where this reduced expression was due to genomic copy number aberrations were eliminated allowing the identification of potential LRES domains (Stransky et al., 2006) (Figure 1.19). Importantly this was the first (and only) study to incorporate a CGH analysis of copy number aberrations into the search for LRES. In all, 28 copy-number independent regions of deregulation were identified and these have been associated with increased histone methylation, both H3K9me3 and H3K27me3, and histone hypoacetylation, without any necessary change in DNA methylation (Stransky et al., 2006; Vallot et al., 2010). Further work showed that regions of LRES can exist in subsets of bladder cancer and that this phenotype correlates with the higher stage and aggressiveness of the cancer (Vallot et al., 2010).

Using a similar computational approach 47 LRES domains were identified in prostate cancer using a sliding window that seeks out windows of "lowly expressed genes"' which were associated with gain of repressive histone and methylation marks (Coolen et al., 2010). Although both of these studies describe the systematic identification of large regions of epigenetic repression in cancer there is no consensus as to how these regions arise apart from altered epigenetic landscapes.

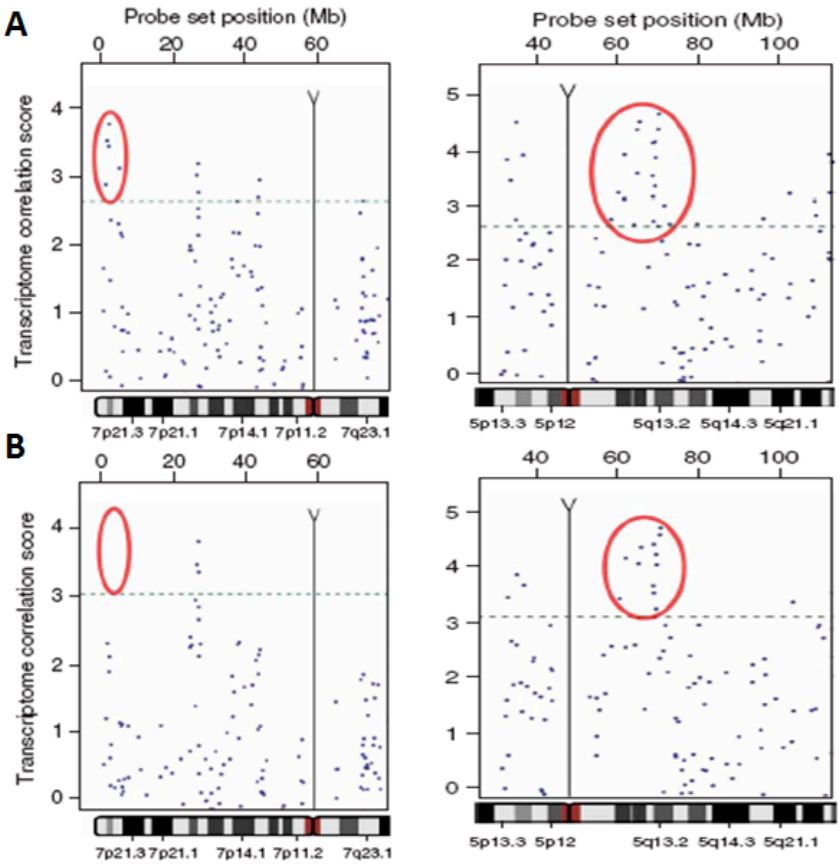


Figure 1.19: Identifying LRES domains in Bladder Carcinoma. Transcriptome Correlation Maps (TCMs) showing scores of how well a gene's expression correlates with seven neighbouring genes before (A) and after (B) removal of genes that are affected by a copy number aberration. Left panel – a copy-number independent region of co-ordinately misregulated genes (circled). Right panel – a copy-number dependent region of co-ordinately misregulated genes (circled) (From Stransky et al., 2006)

By integrative analysis of coordinate gene expression and ER binding data, 11 regions of LRES in breast cancer were identified in association with estrogen signalling and altered chromatin topology as assessed by chromosome conformation capture (3C) (Hsu et al., 2010). It was shown that at one such locus on chromosome 16p11.2 the repression was estrogen-inducible in normal cells and was associated with the formation of a large DNA loop structure that brought together the promoters of the 14 silenced genes in the region (Hsu et al., 2010). Moreover it was found that in breast tumour cells this loop structure was not transient but a permanent feature of the cells indicating a novel mechanism of LRES involving alteration of the chromatin structure itself (Hsu et al., 2010). The existence of a permanent loop structure is contentious given that chromatin fibers are

constantly moving and undergoing nucleosomal rearrangements, however it is possible that the effects seen by Hsu et al due to persistent signalling stimulation of the ER leads to contacts between the gene promoters more frequently than would be expected under normal circumstances.

Prior to my PhD, research in the field of long-range regulation was focussed on aberrant silencing (LRES) in cancer, as discussed in the examples above. At the time of writing this thesis the first example of long-range epigenetic activation (LREA) was published (January 2013). Bert et al (2013) show in prostate cancer, using an integrated genomic approach to uncover regions of multiple upregulated neighbouring genes, that these regions undergo epigenetic activation by chromatin remodelling and DNA methylation changes across the LREA domains. Furthermore the prostate cancer associated LREA regions contained the biomarker genes, oncogenes and microRNAs which were activated in association with a loss of repressive- and a gain of active- histone marks (Bert et al., 2013).

1.4 Aims of Thesis

Identifying domains of coordinate misregulation in breast cancer with a view towards investigating the mechanism in terms of chromatin architecture

The core aim of this thesis is to use publically available datasets to identify regions of misregulation in the breast cancer genome that are independent of genomic amplifications/deletions and could therefore be potential regions of epigenetic regulation (RER). This was to be achieved using expression microarray and CGH array datasets for breast tumours and cell lines to look for genes that fit the following criteria:

1) The genes are positioned local to one another
2) The genes in the region would be co-ordinately expressed
3) The changes in gene expression would be independent of genomic changes that could account for the coordinate gene expression pattern

LRES has become a common phenomenon in the cancer literature and has been attributed to a whole medley of gains/losses of various histone modifications and DNA methylation changes. Given that there is no consensus as to the cause of LRES my hypothesis was that nuclear organization, in particular chromatin architecture has a role in the misregulation of gene expression over such large regions of contiguous genes. Therefore, having identified co-ordinately misregulated regions in breast cancer tumours and cell lines, my aim was to investigate how compact the chromatin at these regions were, how this correlated with the changes in gene expression and the mechanisms underlying this chromatin conformation.

Chapter 2

Materials and Methods

2.1. Reagents

During my PhD I was fortunate to benefit from the excellent Core Scientific Services at the MRC Human Genetics Unit (HGU). Solutions and ingredients provided by these facilities are described below:

L-Broth: Tryptone 10 g, Yeast extract 5 g, NaCl 10 g, Glucose 1 g. Made up to 1 litre with water and autoclaved.

L-Agar: 15 g agar was added to 1 litre L-broth to prepare solid media.

Tris-HCl: Tris[hydroxymethyl]aminomethane (Tris base) was dissolved in sterile water and pH adjusted using HCl.

EDTA: Ethyldiaminetetra-acetic acid di-sodium salt was dissolved in sterile water, solid NaOH added to bring pH to 8.0.

TE buffer: 10 mM Tris.HCl (pH 8.0), 1mM EDTA

PBS (Phosphate buffered saline): NaCl 8g, KCl 0.2g, Na_2HPO_4 1.44g, KH_2PO_4 0.24g , all in a final volume of 1 litre of distilled water with the pH adjusted to 7.4.

SSC (Saline Sodium Citrate) (20X): NaCl 175.3g; Sodium citrate 88.2g – in a final volume of 1 litre of distilled water and pH adjusted to 7.0.

2.2. Cell Culture

2.2.1. Cell lines

Normal breast epithelial and cancer cell lines utilised for the experimental part of this thesis are listed in table 2.1. Culture conditions for all cell lines were 37°C with 5% CO_2. Cell lines were stored at -80°C in fetal calf serum (FCS) supplemented with 10% DMSO (dimethyl sulfoxide) before transfer to liquid nitrogen storage tanks. Cells from liquid nitrogen storage were thawed at 37°C, transferred to their respective culture media and spun down by centrifugation (1000g for 4 mins). The cell pellet was then resuspended in the culture media and plated into either t25cm² or t75cm² tissue culture flasks depending on the number of cells thawed (GIBCO).

Table 2.1. Cell lines and culture reagents.

Cell Line Name	Gene Cluster (Neve et al., 2006)	ER status	Culture media
MCF7	Luminal	positive	DMEM, 10% FCS, 1% P/S
LY2	Luminal	positive	DMEM, 10% FCS, 1% P/S
MDAMB231	Basal B	negative	DMEM, 10% FCS, 1% P/S
MDAMB468	Basal A	negative	L-15, 10%, 1% P/S
MDAMB361	Luminal	positive	DMEM, 10% FCS, 1% P/S
HMLE	NA	NA	MEBM, 1% P/S

FCS, fetal calf serum; P/S, Penicillin (10,000 units/ml)/Streptomycin (650µg/ml), DMEM, Dulbecco's modified Eagle's medium (GIBCO #11965-092); RPMI, RPMI medium 1640 (GIBCO #27016-021); Ham's F12, F-12 nutrient mixture (Ham) (GIBCO #11039-021); L15, Leibovitz's L-15 medium (GIBCO #11415-064); MEGM, Mammary Epithelial Growth Media (Lonza #CC-3151).

Cells were grown to a semi-confluent state (as assessed by light microscropy at 40x magnification) before being harvested and transferred into new culture flasks to ensure their continued growth. To passage adherent cells, all culture media was removed by

careful aspiration and then the cells washed with PBS (phosphate buffered saline solution). Trypsin-EDTA (Sigma) (10%) was added and the flask incubated at 37°C for 3-5 minutes depending on the cell line. Cells were dislodged by agitation of the flask and combined with culture medium (90%) to inactivate the trypsin-EDTA. Cells were pelleted by centrifugation (1000g for 4 minutes) and resuspended in culture media for plating in new culture flasks. A confluent flask of cells was typically split 1:3.

2.2.2. Synthesis of Stripped Fetal Calf Serum

For hormone deprivation experiments FCS was stripped of all endogenous steroids using the following protocol obtained from Ben Skerry (Edinburgh Cancer Research Centre). Fetal calf serum (1 litre) was heat inactivated in a waterbath set at 56°C for 30 minutes before adding 2000U/l of sulphatase (stock 22400U/g therefore 0.089g/litre needed). The serum was incubated with the sulphatase for a further 2 hours at 37°C and then the pH adjusted to 4.2 using HCl. A charcoal mix (for 1 litre: 5g charcoal, 25mg dextran T70, 50ml water) was then added and incubated overnight at 4°C with a magnetic pellet stirring the mixture. The following day the charcoal was removed by centrifugation at 500x g for 30mins at 4°C. The pH was then re-adjusted to 4.2 and a second charcoal mix (as above) added. This was incubated as before at 4°C overnight with magnetic stirring. The next day this second charcoal mix was removed by centrifugation at 10,000 rpm for 30mins at 4°C. This centrifugation step was repeated to remove any residual charcoal and the pH adjusted to 7.2 with NaOH. Stripped FCS was filter sterilised, aliquoted into 50ml falcon tubes and stored at -20°C.

2.2.3. Estrogen treatment of cells in culture

Semi-confluent cell cultures were transferred into phenol-free DMEM (Gibco) supplemented with 5% L-Glutamine, 5% P/S and 10% stripped FCS. Cell cultures were kept under "starvation" conditions in this media for 72 hours in a cell culture incubator at 37°C and 5% CO_2. Cultures were then induced with 100nM of 17ß-estradiol (Sigma) for 0-24 hours before harvesting.

2.3. Fluorescence in-situ hybridization (FISH)

2.3.1. Preparing FISH probes

Genomic clones used for FISH

The UCSC browser was used to identify genomic clones flanking regions of interest for FISH experiments. Fosmids were obtained from the BACPAC Resources Centre at the Children's Hospital Oakland Research Institute (http://bacpac.chori.org/). All fosmid clones were provided as bacterial stab cultures in agar. A full list of all fosmids used in this thesis are listed in Table 2.2 and the regions for which probe pairs were used in Table 2.3.

Table 2.2 Names and Genomic Positions of Fosmids used in this thesis. All coordinates are based on hg19/Feb 2009 release of the UCSC genome browser (http://genome.ucsc.edu/).

Name			Coordinates		
UCSC browser ID	BACPAC whitehead ID	Chromosome band	Start	End	Size (bp)
G248P86075G2	WI2-1584N4	16p11.2	29664946	29703636	38690
G248P86565D5	W12-1754H9	16p11.2	30010045	30046946	36901
G248P8190D5	W12-906G10	16p11.2	30379725	30424227	44502
G248P8178C9	W12-497E18	16p11.2	30788769	30824965	36196
G248P87014G2	W12-2222M4	16p11.2	31166763	31207277	40514
G248P85537E3	W12-2033J5	11p15.4	5453348	5494460	41112
G248P8086G3	W12-528M6	11p15.4	4961240	4999789	38549
G248P89117G5	WI2-2889N9	16p11.2	28899528	28938582	39054
G248P89434H1	WI2-3081O2	16p11.2	28483429	28519288	35859

Table 2.3 Fosmids used as probes for each FISH locus with coverage (kb).

Region	Chromosome band	5' Probe	3' Probe	Locus Size (kb)
RER locus 1	16p11.2	WI2-1584N4	W12-1754H9	382.00
RER locus 2	16p11.2	W12-1754H9	W12-906G10	414.18
RER locus 3	16p11.2	W12-906G10	W12-497E18	445.24
RER locus 4	16p11.2	W12-497E18	W12-2222M4	418.51
Control locus	11p15.4	W12-2033J5	W12-528M6	453.55
Control locus	16p11.2	WI2-2889N9	WI2-3081O2	380.24

Bacterial culture and stocks

Fosmids were streaked out onto LB-agar plates containing chloramphenicol (25µg/ml - made up in Ethanol) and incubated overnight at 37°C. Single colonies were picked using a sterile pipette tip which was added to a 5ml L-Broth supplemented with chloramphenicol (25µg/ml). These were grown overnight in a 37°C incubator with shaking. Bacterial stocks were made of all fosmid clones taking a 1:1 mixture of the overnight culture of L-Broth:Glycerol. This mixture was stored at -70°C.

Fosmid minipreps

L-Broth cultures were pelleted by centrifugation at 16,000g for 30 seconds and resuspended in 200 µl GTE buffer (50mM glucose, 25 mM Tris pH 8, 10 mM EDTA) containing freshly added lyzozyme (Sigma) for 5 minutes at room temperature. To this 400µl of ice-cold lysis buffer (0.2 M NaOH, 1% SDS) were added and mixed by inversion before a further 5 minute incubation on ice. This was followed by the addition of 300µl acetate buffer (5M potassium acetate, 11.5% glacial acetic acid) to precipitate out cell debris by 5 minute incubation on ice (white flocculent precipitate formed in tube). The precipitate was pelleted by centriguation at 16,000g for 5 minute at 4°C and the supernatant was combined with an equal volume of phenol:chloroform. Following centrifugation at 4°C for 2min at 16,000*g* the aqueous top layer was removed to a fresh tube and mixed with an equal volume of isopropanol and incubated at -20°C for at least one hour. The DNA was pelleted by 15min centrifugation at 16,000*g* at 4oC, washed in 70% ethanol and air-dried before resuspension in 20-30µl of TE (depending on the pellet

size). Rnase A (1mg) was added to the resuspended DNA and incubated for 5 minutes at 37°C. All fosmid DNA preparations were stored at -20°C.

Labelling of probes by nick translation

DNA probes for FISH were labeled by nick translation to incorporate either biotin-16-dUTP or digoxigenin-11-dUTP (Roche). To do this 1-1.5 µg DNA were added to 4µl nick translation salts (0.5 M Tris-HCl pH 7.5, 0.1M $MgSO_4$, 1mM DT, 500 µg/ml BSA) with 5 µl of each of 0.5 mM dATP, dCTP and dGTP. Then either 5 µl of 1mM biotin-16-dUTP or 3 µl of 1mM digoxigenin-11-dUTP with 2 µl of 0.5mM dTTP were added to the mixture. DNaseI (Invitrogen) was diluted 1:500 and added to the reaction to a final concentration of 1U/ml. The reaction was made up to 40 µl with water and 1 µl of DNA polymerase I was added (10U/µl). After mixing the nick translation reaction ws left to proceed at 16°C for 90 minutes. Nick translation was stopped by the addition of 3µl of EDTA and 2µl of 20% SDS. Unincorporated nucleotides and enzymes were removed using Quick Spin Columns (Roche) as per the manufacturer's instructions and eluting the labeled probes with TE pH7.5. Labelled probes were quantified (as below) and stored at -20°C.

Quantitation of labeled FISH probes

Labelled DNA probes for FISH were spotted onto a nitrocellulose membrane which had been prepared by soaking in water, followed by 20xSSC for 10 minutes and allowed to dry. Probes were spotted in dilutions of 1:500, 1:1000, 1:5000 and 1:10000 and allowed to dry on the membrane. Standards of known concentration for biotin and digoxigenin (Roche) were also spotted on the membrane at 20, 10, 2 and 1pg. Once dry, the DNA was cross-linked to the membrane by exposure to UV irradiation (150mJ). The membrane was then soaked briefly in buffer 1 (0.1M Tris-HCl pH 7.5, 0.15M NaCl) at room temperature before being immersed in block solution (buffer 1 supplemented with 3% BSA) at 60°C for 60 mins. The blocked membrane was then incubated at room temperature (with gentle agitation) with strepavidin-alkaline phosphatase and anti-digoxigenin-alkaline phosphatase (diluted 1:1000, Roche). The membrane was then washed twice in buffer for 15 mins at room temperature with agitation, then incubated in a sealed polythene bag with 5ml of buffer 3 (0.1M Tris-HCl, pH 9.5) and 2 drops from

each solution in the alkaline phosphatase substrate kit VI (Vector laboratories). A blue reaction product is formed by 5-bromo-4-chloro-3-idolyl phosphate and nitoblue tetrazoilium and allows estimation of the concentration of labeled DNA by comparison with the standards of known concentration.

2.3.2. Preparing cells for 2D FISH

KaryoMAX Colcemid (0.1 μg/ml) was added to cell culture 30mins prior to harvest to arrest the cells in metaphase through inactivation of spindle microtubules. This is important to increase the number of metaphase chromosome spreads in the preparation for karyotyping the cells (when confirming probe localization see section 2.4.8). Cells were then rinsed in PBS and resuspended in hypotonic solution (75mM KCl). Hypotonic solution was added gradually (drop-wise) with constant vortex agitation to avoid the formation of clumps keeping a single cell suspension. Cells were left to swell in hypotonic solution for 10 minutes at room temperature before centrifugation for 5 minutes at 400g and aspirating off the hypotonic solution. The nuclear pellet was then resuspended in MAA fix (3:1 methanol:acetic acid) in a drop-wise manner with constant agitation. Fix was then removed by centrifugation and the pellet resuspended in fresh MAA fix to be stored at -20°C or used immediately for preparation of FISH slides.

2.3.3. Preparation of slides for 2D FISH

Glass slides that had been soaked in ethanol (with a few drops of concentrated HCl) for at least an hour were cleaned and dried for use. The fixed cell preparations were centrifuged for 5 minutes at 1000g and fresh fixing solution added (a few drops or until the suspension was slightly “milky” in appearance). From a height of approximately 30 cm (“arm’s-length”) a drop of this suspension was dropped using a fine-tipped pastette onto the glass slide. Quality of the spread was improved by humidities of around 50% and if the cells were moist (by breathing on the slide prior to dropping). Slides were kept for 2-7 days before being used for hybridization or artificially aged by baking the slides at 60°C for 1 hour prior to beginning FISH.

2.3.4. Preparation of tissue sections for 3D FISH

Parrafin-embedded tissue sections were obtained with help from Jeremy Thomas (Consultant Pathologist, The Western General Hospital). Usage of tumour material was approved by the Lothian Research Ethics Committee (08/S1101/41) and obtained under the auspices of Experimental Cancer Medicine Centre programme (Edinburgh). Tissue sections for FISH were cut by Lynne Johnstone (Welcome Trust Clinical Research Facility) at 6μm and laid on Superfrost+ slides.

The slides were baked at 65°C for 30 minutes to melt the wax, washed four times in 200ml xylene for 10 minutes, rehydrated through an ethanol series (4x 10 minute washes in each of 100%, 95% and 70% ethanol) before being microwaved for a further 30 minutes in 0.1M citrate buffer (pH 6). The slides were then allowed to cool for 20 minutes in the citrate buffer solution before being washed and stored in water. Slides were rinsed in 2x SCC before use.

2.3.5. Hybridisation of labeled FISH probes

For 2D FISH, slides were incubated for 1 hour at 37°C with RNaseA (120 μg/ml) in 2 x SSC, rinsed in 2 x SSC at room temperature and then dehydrated through a series of 2 minute ethanol washes (70%, 90% then 100%) and left to air dry. Slides were then heated in a 60°C oven for 5 minutes before being transferred to denature solution (70% formamide in 2 x SCC pH7.5) at 70°C for 1 minute. After denaturation slides were then transferred to 70% ethanol on ice for 2 minutes before further dehydration in 90% and 100% ethanol at room temperature.

For 3D FISH slides were washed in 2 x SSC at 75°C for 5 minutes then denatured for 3minutes at 75°C in 70% formamide/2xSCC pH7.5. Slides were then placed in ice cold 100% ethanol for 3 minutes before further dehydration in 90% and 100% ethanol at room temperature (as for 2D FISH).

The FISH probes were prepared for hybridization by combining 100-150ng of labeled DNA with 5 μg salmon sperm DNA per slide and 12μg of human Cot1 (supplier) DNA per probe. To this 2 x volume ethanol was added and the probes spun under high heat (in

a lyopholiser) to produce a white pellet of DNA which was resuspended in 25µl of freshly prepared hybridization mix (50% deionised formamide, 10% dextran sulphate and 1%Tween 20 in 2 X SSC). Probes were incubated at 70°C for 5 minutes to allow denaturation and then 37°C for 15 minutes so they could reanneal before being pipetted onto 22 x 22mm coverslips (on a hot plate) and picked up using the slides. The slides were sealed using TipTop rubber solution around the coverslip and incubated in a 37°C waterbath in a covered tray overnight (or 16 hours).

2.3.6. Washing and detection of FISH probes

After overnight hybridization, the rubber solution sealing the slides was removed and the slides were washed 4 x in 2 x SSC for 3 minutes at 45°C. Coverslips fell off in the solution during the first wash and if they did not were gently agitated in the solution until they did. Slides were then washed 4x in 0.1 x SSC for 3 minutes at 60°C and transferred to 4 x SSC/0.1% Tween 20 at room temperature.

Slides were incubated for 5 minutes with blocking buffer (4X SSC, 5% Marvel) before antibody detection. All FISH antibodies were diluted in 5% Marvel in 4 X SSC. Digoxigenin-labelled probes were detected using sequential layers of Fluorescein isothiocyanate (FITC)-conjugated anti-digoxygenin (200ug/ml stock diluted 1:20) and FITC-conjugated anti-sheep IgG (1.5mg/ml diluted 1:100). Biotin-labelled probes were detected with sequential layers of Texas Red (TR)-conjugated avidin (1mg/ml stock diluted 1:500), biotinylated anti-avidin (BAA) (0.5mg/ul diluted 1:100) and TR-conjugated avidin (1mg/ul diluted 1:500) (Suppliers for all of these reagents). All antibodies were obtained from Vector Laboratories with the exception of FITC-conjugated anti-digoxygenin which was obtained from Roche. Detection was carried out in a humidity chamber for 30 minutes which was kept at 37°C. Slides were washed 3x after each antibody-incubation for 2 mins (with agitation) in 4 X SSC/0.1% Tween at 37°C before addition for the next antibody layer.

Slides for 2D FISH were mounted in Vectashield (Vector) with 0.5µg/ml DAPI. Slides for 3D FISH were incubated in 4 X SSC/1% Tween with 50ng/ml DAPI for 5 minutes

before mounting in Vectashield (Vector). Coverslips on all slides were sealed in place with rubber solution.

2.3.7. Image capture

Examination of nuclei after 2D FISH was carried out using a Hamamatsu Orca AG CCD camera (Hamamatsu Photonics (UK) Ltd, Welwyn Garden City, UK) fitted to a Zeiss Axioplan II microscope with Plan-neofluar oil-immersion objectives, a 100 W Hg source and Chroma #8300 triple band pass filter set. Image capture and analysis of nuclear size and distance between the hybridization signals was performed with scripts written for IPLab Spectrum (Scanalytics Copr, Fairfax, VA) as described in section 2.4.10.

Examination of nuclei from tissue sections by 3D FISH was carried out using a Hamamatsu Orca AG CCD camera (Hamamatsu Photonics (UK) Ltd, Welwyn Garden City, UK), Zeiss Axioplan II fluorescence microscope with Plan-neofluar or Plan apochromat objectives, a Lumen 200W metal halide light source (Prior Scientific Instruments, Cambridge, UK) and Chroma #89014ET single excitation and emission filters (Chroma Technology Corp., Rockingham, VT) with the excitation and emission filters installed in Prior motorised filter wheels. A piezoelectrically driven objective mount (PIFOC model P-721, Physik Instrumente GmbH & Co, Karlsruhe) was used to control movement in the z dimension. Hardware control, image capture and analysis were performed using Volocity (Perkinelmer Inc, Waltham, MA). Images were captured at 200 nm intervals in the z axis and were deconvolved using a calculated PSF with the constrained iterative algorithm of Volocity. Image analysis was carried out using the Quantitation package of Velocity as described in section 2.4.10.

2.3.8. Confirmation of probe localisation by FISH on metaphase preparations

The genomic localisation of all probes was verified by 2D FISH of metaphase spread preparations from cells with normal karyotype. All MAA fixed cell preparations for this analysis were obtained from Shelagh Boyle. Karyotyping and confirmation of probe localisation to the correct chromosome was carried out with the aid of expertise in this area from Shelagh Boyle and Elizabeth Kerr. An example of this, for RER locus 2, is shown in figure 2.1.

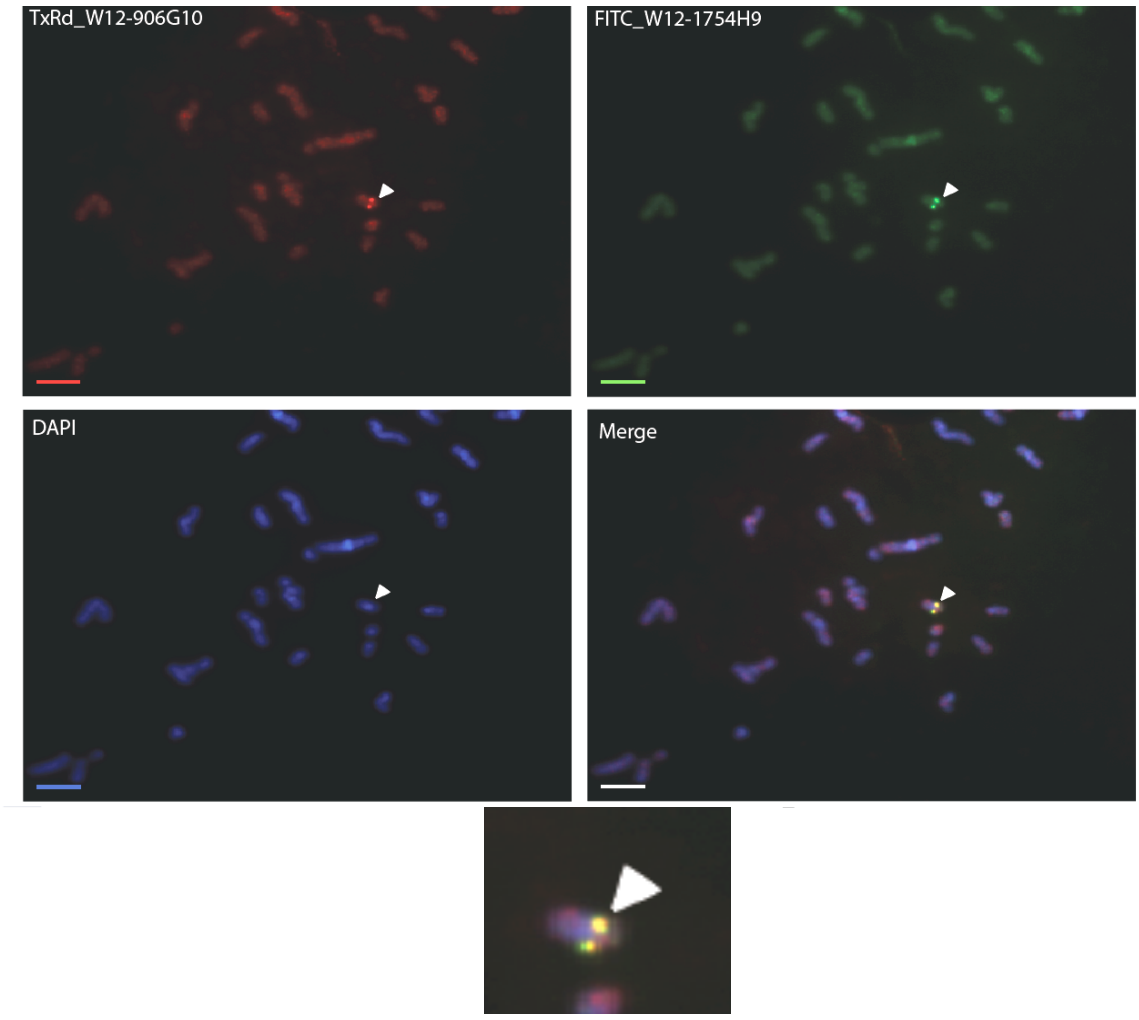


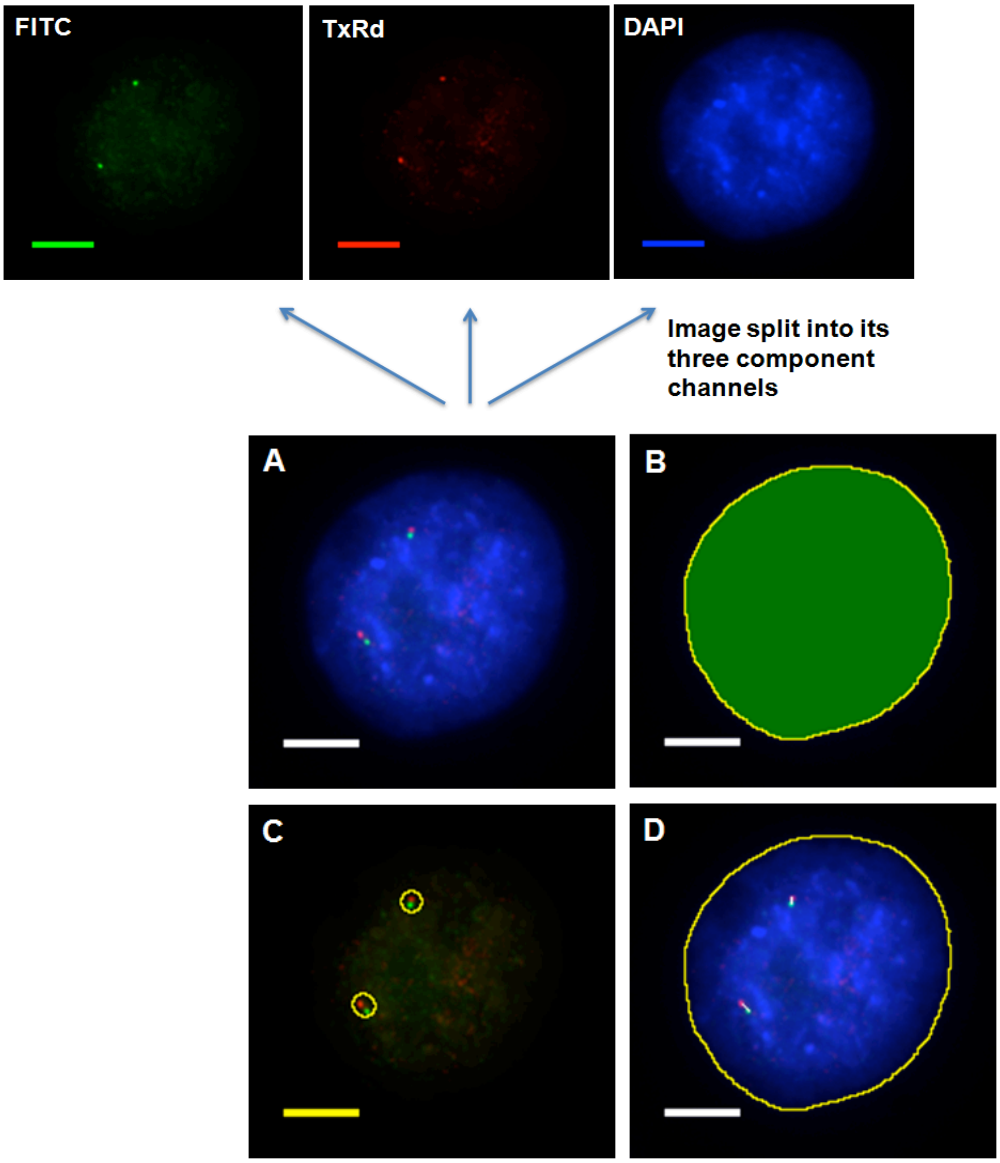
Figure 2.1. Confirmation of FISH probe localisation. Metaphase spreads from human blood fibroblast cells (obtained from Shelagh Boyle) were hybridised with digoxigenin- and biotin- labelled probes by FISH to verify their chromosomal localisation. Shown are the probes for RER locus 2 with white arrows indicative of their position on chromosome 16p11.2.

2.3.9. Scripts for image analysis

All scripts for image analysis of 2D FISH images were written for IPLab by Paul Perry (Microscopy & Imaging Department). Scripts for analysis of 3D FISH images in Velocity were written by Shelagh Boyle (Research Assistant, Bickmore Lab) using packages set up for Velocity by Matthew Pearson (Microscopy & Imaging Department).

To calculate inter-probe distances in 2D, between 50-60 of the most evenly shaped nuclei with clear pairs of hybridisation signals were selected by systematic scanning of slides by eye. For each probe the IPLab script calculated the pixel in the image with the highest intensity. The highest intensity pixel in the centroid for each probe in a pair were selected and the distance between them calculated. For analysis of flattened 3D images from tissue section FISH the inter-probe distances were calculated by taking the centroid of each probe in the pair and calculating the distances between them. All inter-probe distances from 2D and 3D FISH were converted from pixels to microns by multiplification by 0.134μm. The 2D images were also segmented by DAPI staining to calculate the nuclear area (pixels). This measurement was used to calculate the nuclear radius which was also converted to μm units of measurement. This analysis is illustrated in Figure 2.2.

For 2D FISH data, inter-probe distances (d) and radii (r) were then squared and normalised to account for differences in nuclear size. Therefore all normalised inter-probe distances described in this thesis are ratios of d^2/r^2. The difference between the distribution of squared inter-probe distances between datasets was assessed statistically using the Wilcox test with a cut off of $p<0.05$.



E

	Area (pixel)	Distance (pixels)
Pair 1	15061.00	2.91
Pair 2	15061.00	4.30

Figure 2.2: Measuring the inter-probe distance and nuclear area from FISH analysis. The script for image analysis was written for IPLab by Paul Perry. Shown is an example of a DAPI stained nucleus (blue channel) with FISH signals (red and green channels) representing the locus of interest (A). The image was segmented by the intensity of DAPI staining and the nuclear area estimated by the number of pixels within the segment (B). The image was then split to show the red (TxRd) and green (FITC) channels only and a paint tool was used to manually select the pairs of probes (C). The signal intensities for the red and green probes were calculated. The pixel with the highest signal intensity for each probe was to calculate the distance between the two probes (D). The script generated a data table for each image giving the nuclear area and interprobe distances in pixels.

To determine the radial positioning of probe pairs a nuclear erosion script was used in IPLab for analysis of 2D FISH images. This script eroded the nuclei into five concentric shells of equal area as illustrated in Figure 2.3. It was then possible to count the distribution of probe hybridisation signals across the erosion shells in 50-60 nuclei. The number of hybridisation signals in each shell in the dataset were calculated as a percentage of the total number of signals and plotted as a histogram. Significance of changes in radial positioning of probes between cell lines was assessed using the fisher test with a cut off of $p<0.05$.

This analysis was not carried on imaging data from 3D FISH of tissue sections as the nuclear erosion script is limited to analysis of single nuclei of regular shape that are not touching any other nuclei or stained structures.

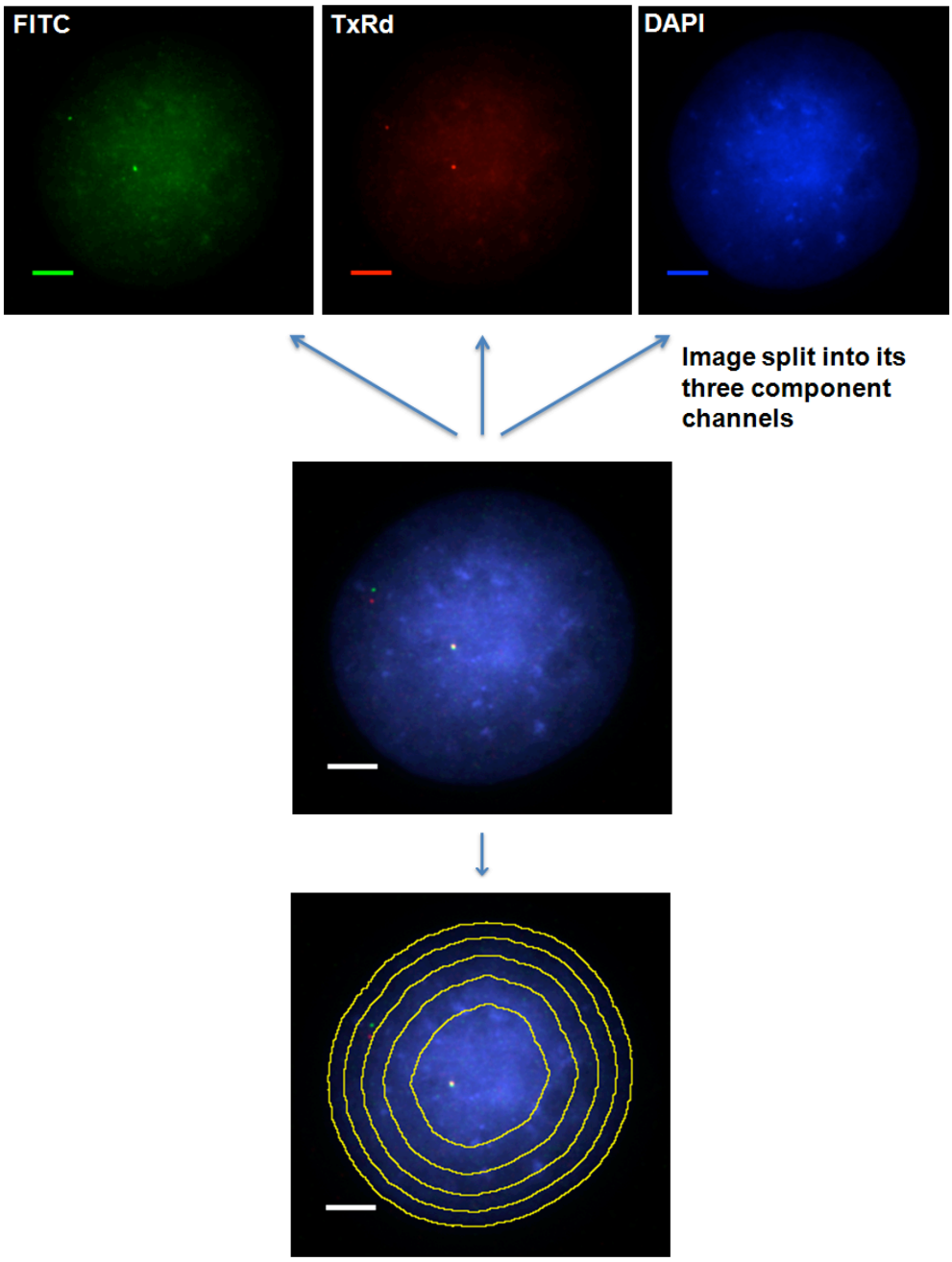


Figure 2.3: Measuring the probe localisation by nuclear erosion of FISH images. The script for image analysis was written for IPLab by Paul Perry. Shown is an example of a DAPI stained nucleus (blue channel) with FISH signals (green and red). The DAPI segment of the image was eroded into five consecutive shells of equal nuclear area (as calculated by the number of pixels in each shell). It was then possible to manually tally how many probe signals were in each shell.

2.4. Computational Biology

A number of websites were used for computational components of the work compiled in this thesis:

Ensembl	http://www.ensembl.org/
R-Project	http://www.r-project.org/
Bioconductor	http://www.bioconductor.org/
Gorilla	http://cbl-gorilla.cs.technion.ac.il/
UCSC Browser	http://ucsc.genome.edu/
OMIM	http://ww.ncbi.nlm.nih.gov/omim
GEO	http://ww.ncbi.nlm.nih.gov/geo
Circos	http://www.circos.ca/
Tinn R	http://www.sciviews.org/Tinn-R/index.html

Basic quantitative analysis of data was carried out in Excel spreadsheets (Microsoft) using default settings.

All statistical analysis was carried out in R open source statistical language (mainly version 2.15.1 and 2.15.2).

Bar charts were drawn in Excel (Microsoft). Circular visualization of data was carried out in Circos. Boxplots were drawn in R where the median of the data was expressed as a heavy line and the boxed area representing the interquartile range (outliers were removed). Heat maps were plotted using the image function in R. TCS maps were plotted in R with a heatmap image representing the chromosome ideogram (coordinates obtained from Duncan Sproul).

A large component of the work compiled in this thesis (chapters 3 & 6) was based on computational analysis of publically available datasets in R. The methods used for these analyses are described in chapter 3 of this thesis.

Chapter 3

Identifying Regions of Epigenetic Regulation (RER) in Breast Cancer

3.1: Introduction

Large scale gene expression profiling studies have helped our understanding of the clinical and pathological differences in breast cancer, leading to the conclusion that it is not a single disease but a collection of molecularly heterogeneous disorders (Sørlie, 2004; Sotiriou and Pusztai, 2009). Whilst genetic aberrations which cause misregulation of genes and genomic stability are a hallmark of cancer, epigenetic changes also have crucial influences on tumourigenesis and cancer treatment and could therefore be expected to contribute to breast tumour heterogeneity (Jones and Baylin, 2007; Kristensen et al., 2009).

Long range epigenetic silencing (LRES) has become a common term to help explain some of the aberrant gene expression seen in cancer (Coolen et al., 2010; Dallosso et al., 2012, 2009; Frigola et al., 2006; Hsu et al., 2010; Javierre et al., 2011; Novak et al., 2008, 2006; Park et al., 2011; Stransky et al., 2006; Vallot et al., 2010). Importantly the LRES status has been correlated with aggressiveness of tumours and therefore has implications for patient prognoses (Dallosso et al., 2009; Vallot et al., 2010). Therefore, where previous classification of tumours has been based on traditional methods like histopathological grading and staging with newer prognostic predictions based on molecular signatures, finally an epigenetic classification is emerging which describes the link between the misregulation of genes in cancer and disease pathways.

I set out to systematically identify large contiguous regions of the breast cancer genome that were misregulated, potentially by epigenetic mechanisms. I used a computational approach akin to that used previously in the study of breast and bladder cancer (Reyal et al., 2005; Stransky et al., 2006). My intention was to mine publically available breast cancer datasets for which both gene expression and copy number data were available. The latter was important to eliminate copy number gain or loss as causative of increased or decreased levels of gene expression.

3.2: Methods

3.2.1: Datasets used

Stransky et al. 2006. Affymetrix Human GenomeU95A & U95Av2 arrays consisting of 12,500 probes were used for analysis of gene expression on 57 bladder carcinomas. DNAs were also analyzed on CGH microarrays (HumArray 2.0) containing 2,385 BAC clones at 1.3 Mb intervals. Normalized expression data was obtained from ArrayExpress (E-TABM-147) and normalised CGH from http://microarrays.curie.fr/publications/

oncologie_moleculaire/bladder_TCM/. (Stransky et al., 2006)

Jöhnsson et al., 2010. Oligonucleotide arrays (Gene Expression Omnibus, GEO, platform GPL5345) and BAC microarrays (GEO platform GPL4723) consisting of 32,000 clones were used for global analysis of gene expression and copy number in 359 breast tumours (Jönsson et al., 2010). I used expression and copy number data for 356 tumours common to both datasets. All gene identifiers were mapped to Ensembl annotations using Ensembl BioMart. Where multiple probes were mapped to a gene the probe with the highest median expression was retained and the remaining multiple probes discarded.

Chin et al., 2006. Copy number profiles for 145 primary breast tumours using Scanning (2464 BACs at 1Mb intervals) and OncoBAC arrays (960 P1, PAC, or BAC clones). Gene expression profiling was available for 130 breast tumours (Affymetrix U133A arrays) (Chin et al., 2006). Expression and copy number data for 96 tumours common to both datasets were used for analysis here.

Chen et al., 2010. Expression data for 42 invasive ductal carcinoma (IDCs) samples and 143 breast tissue samples with normal histopathology were obtained using Affymetrix U133Plus 2.0 GeneChips (Chen et al., 2009). Raw data was processed using standardized RMA normalisation (Bioconductor Affy package) and mapped to Ensembl gene annotation.

Neve et al., 2006. Expression profiles for 51 breast cancer cell lines were obtained using Affymetrix U133A arrays and copy number data using Scanning (2464 BACs at 1Mb intervals) and OncoBAC (1860 P1, PAC, or BAC clones) arrays(Richard M Neve et al., 2006). All expression and CGH data were obtained from the Lawrence Berkeley Laboratories website (http://cancer.lbl. gov/breastcancer/data.php). Expression and copy number data for 48 cell lines common to both datasets were used for analysis here. Raw data was processed using RMA normalisation (Bioconductor Affy package) and mapped to Ensembl gene annotation.

3.2.2: Removing genes affected by copy number aberrations

In order to detect chromosomal copy number aberrations at high resolution, array comparative genomic hybridisation (aCGH) is used. The test DNA sample is hybridised with normal DNA (two copies of autosomal chromosomes) on arrays producing "clones" at base pair positions along the chromosomes. This produces log2 ratios which can be converted to "calls" for copy number gain (e.g. +1, +2, etc), loss (-1, -2, etc) or no change (0). I used the 'R' package CGHcall (van de Wiel et al., 2007) which is available via Bioconductor. This algorithm combines features of previously developed methods (e.g. chromosome breakpoint and segmentation information) to obtain accurate "calls" for the copy number status of each clone.

Genome coordinates for aCGH data clones were mapped to UCSC Human Genome browser (http://genome.ucsc.edu/) build GRCh37/hg19 assembly. Clones were then mapped to their nearest Ensembl gene in the expression data for the respective tumours or cell lines. Copy-number affected genes (gain/loss) in the respective samples were removed from the corresponding expression data for those samples to produce a copy-number independent gene expression file. This data file was used for all subsequent analysis.

Example CGH profiles (plotted as log2 ratios) with the copy number status determined by CGHcall denoted by colour, are shown in Figure 3.1 for three breast cancer cell lines from Neve et al., (2006) and three breast tumours from Jönsson et al., (2010). CGH

profiles for all 48 breast cancer cell lines and 356 tumours are available in supplementary data (Appendices I & II, respectively).

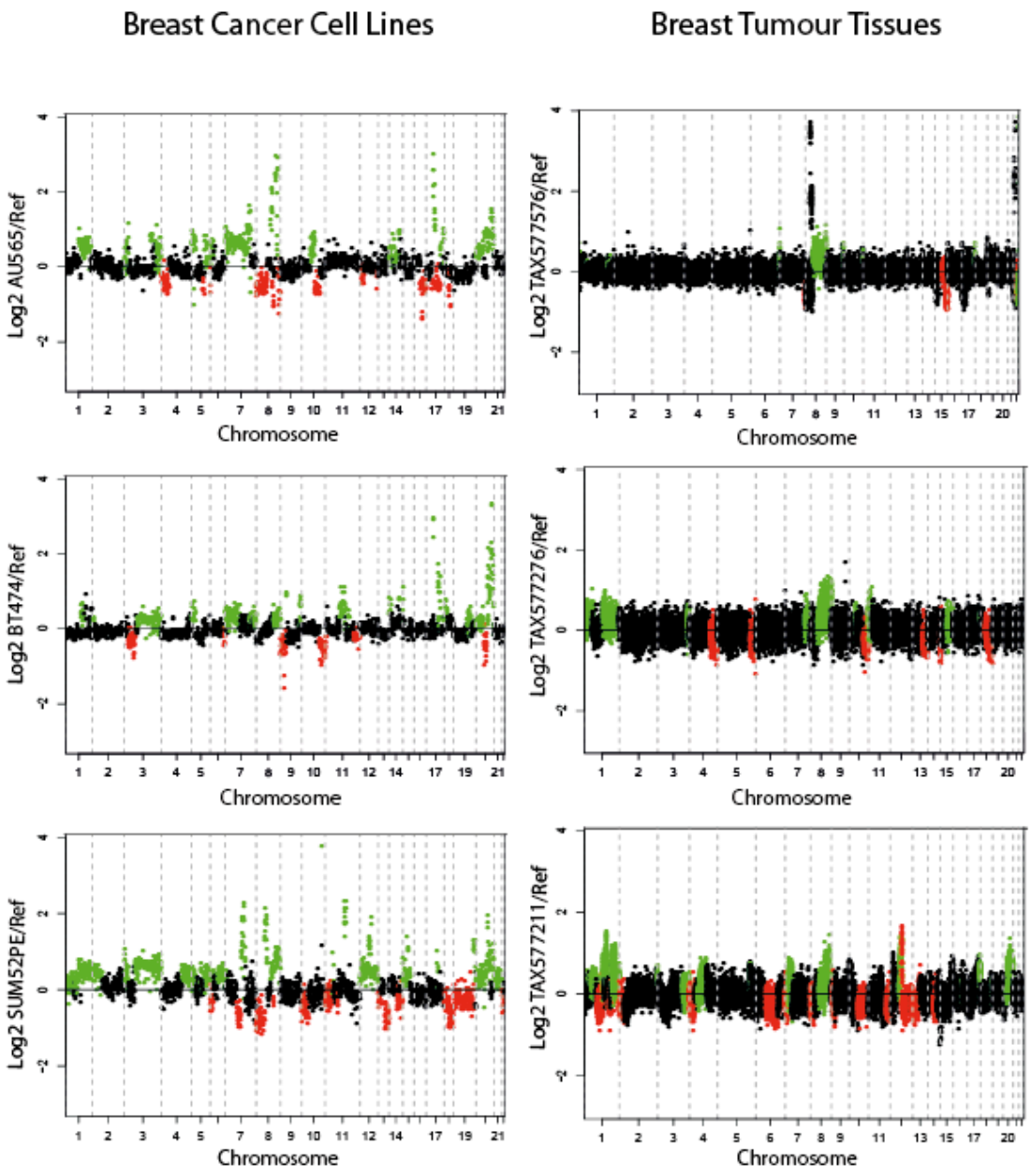


Figure 3.1: DNA Copy Number Profiles in Breast Cancer Cell Lines & Tumours. The log2 (sample:ref) ratios of the clones were plotted against chromosomal position. Regions of loss (green), gain (red), and no change (black) as determined by CGHcall (van de Wiel et al., 2007).

3.3: Results

3.3.1: A sliding window approach to identify genomic regions with coordinate gene expression signatures

An approach for visualising "Transcriptome Correlation Maps" (TCM) which highlight regions of coordinate gene expression patterns across the genome has previously been developed for invasive ductal breast carcinomas (Reyal et al., 2005). This involved calculating, for each gene, the correlation between its expression with that of its neighbours and identified a number of clusters of genes that were coordinately regulated in the tumours. The most significant of these coordinated clusters matched well-established sites of frequently occurring genomic aberrations in breast cancer (e.g. 1q, 8p, 8q, 16p, 16q, 17q and 20q) (Reyal et al., 2005). By later disqualifying regions where the coordinated expression was attributed to a copy number variation, this approach was adapted to identify 28 regions in the genome that are coordinately regulated by long range epigenetic silencing (LRES) in bladder carcinoma (Stransky et al., 2006). LRES regions have also been identified in the prostate cancer genome using a similar approach (Coolen et al., 2010). Whereas this latter method took into account the distance between neighbouring genes, it also selected specifically for regions of low gene expression.

As I was primarily interested in coordinate regulation of neighbouring genes by higher order chromatin structure, regardless of whether it be repressive or activating in nature, I decided the former approach developed by the Radvanyi lab was more appropriate as it accounted for both and was more feasible to assess statistically with a set number of genes in the sliding window.

I adopted a running score method equivalent to that used to find LRES regions in bladder cancer, and applied it to breast cancer. For each gene a transcriptional correlation score (TCS) was calculated which was the sum of the Spearman Rank correlation scores between the RNA levels of each gene with that of each of its neighbouring (*2n*) genes. I

developed an algorithm that would then calculate the TCS for all the genes in the dataset using a sliding window of *2n+1* genes (where *n* is the number of genes either side of the gene of interest). A schematic representation of this sliding window approach is shown in Figure 3.2.

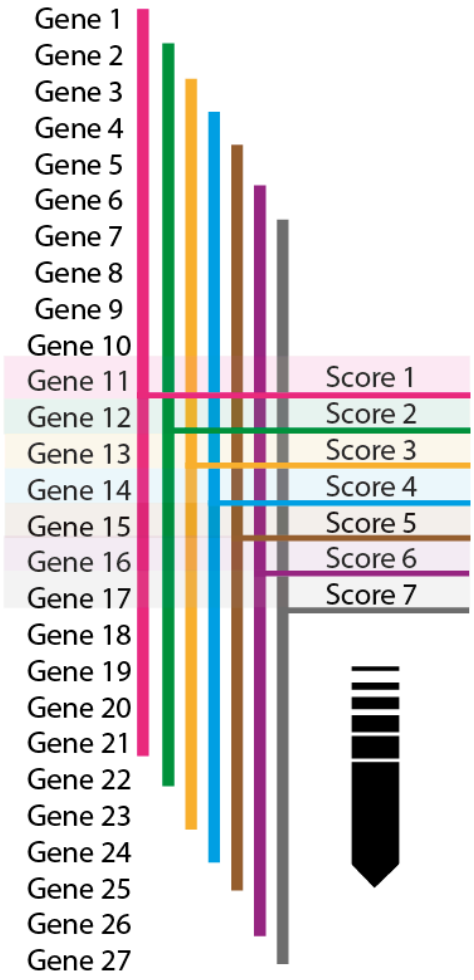


Figure 3.2: Sliding Window Approach to Identifying Regions of Coordinate Expression. For each gene the Spearman Rank Correlation Score is calculated for the expression level of that gene with each of its neighbours across all the samples in a dataset. The scores for these genes are then summed together to get the overall score for each gene. A sliding window approach is then used to calculate the scores for all the genes with their neighbours.

3.3.2 Assessing Significance & delineation of regions

Having developed a method for assigning TCSs to genes in cancer expression datasets, I next wanted to use these scores to identify regions of co-ordinate gene expression in breast cancer. These would be expected to be composed of multiple contiguous genes with high TCS scores. As shown in the example in Figure 3.3 visual inspection of the TCS data using positional maps of the chromosomes shows that there are regions containing clusters of genes with high TCS scores, which form distinct peaks in the maps.

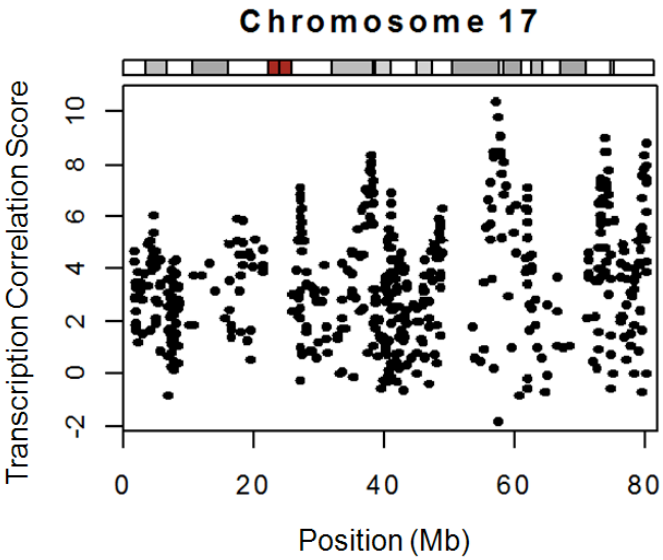
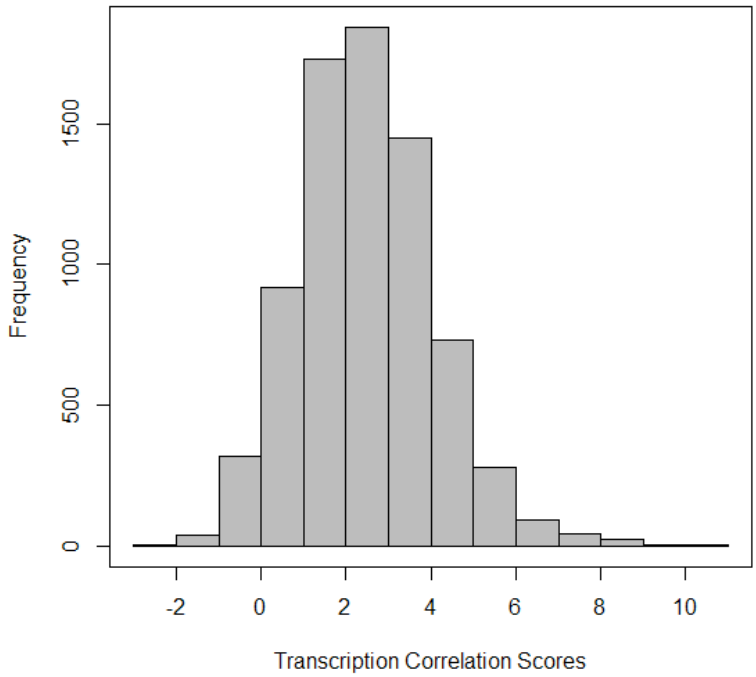


Figure 3.3: Map of transcription correlation for chromosome 17 for 356 breast tumours. Plotted are the transcription correlation scores (TCSs) for each gene, representing the degree of correlation between the expression profile or each gene with its neighbouring genes.

Next I examined the distribution of the TCSs using a histogram which showed that the data resembled a normal distribution with a long positive tail (Figure 3.4A). This suggests that there is random variation in the TCSs but that there are some genes which have elevated scores. To further examine the distribution of the TCS data I plotted a normal Quantile-Quantile (Q-Q) plot (Figure 3.4B). This showed a strong concordance between the TCS scores and a normal distribution but had a tail of outliers at the high end. This confirmed my visual observation from the histogram (Figure 3.4A) and chromosome map (Figure 3.3).

A



B

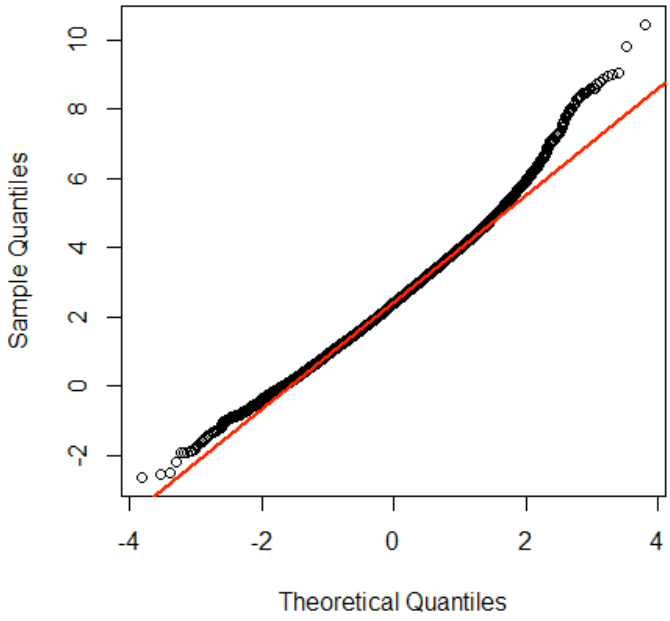


Figure 3.4: TCS data distribution. (A) histogram showing the distribution of all the calculated transcription correlation scores (TCS) across samples (B) Normal Q-Q plot of the calculated TCSs across samples The Q-Q plot was used to plot the quantiles of calculated correlation coefficient against that derived from a normal distribution. Points deviating from the red line indicate deviation from normality.

As a quantitative assessment I tested the normality of the data using the Kolmogorov-Smirnov test which confirmed this deviation (p-value < 2.2e-16).

I therefore identified regions of interest by using z-scores to identify outliers and positional information to delineate regions containing multiple outliers. To do this I used the mean and standard deviations of the TCS score distribution to calculate a z-score for each gene. This was then used to calculate a p-value for each gene based on the probability of observing this z-score in a normal distribution. A p-value threshold was set at $p<0.05$ to determine which genes showed significantly high TCS indicative of coordinated gene expression signatures.

Regions were then delineated by extracting *n* number of genes either side of the significant TCS genes (to get all the genes in the sliding window). Regions containing < 2 significant TCS genes were discarded and overlapping regions were merged together. The regions were then further refined by calculating the median value of gene expression for each gene across all samples in the dataset and working out the correlation between this median for each gene with the rest of the region. It was then possible to assign p-values (Spearman Rank test) for how well correlated each gene was with the rest of the region. The regions were thereby "trimmed" to the first and last gene that had a p value <0.05.

3.3.3: Verifying Method with Published Literature.

I validated my algorithm for detecting regions of coordinate regulation by comparison to the published bladder carcinoma study (Stransky et al., 2006). I used my sliding window algorithm with n=7 (2n+1=15 gene window) neighbouring genes as in Stransky et al. (2006). Figure 3.5 compares the transcriptional correlation map of chromosome 16 from the Stransky et al (2006) supplementary online material (Figure 3.5A: upper panel) and a map generated by my algorithm (Figure 3.5A: lower panel). These data show that my algorithm recapitulated the previously published results.

As I was primarily interested in copy number independent coordinate misregulation I used copy-number independent gene expression data to determine where the regions of coordinate expression/repression were located (file created as described in section 3.2.2.). A significance threshold was established using z scores (cut off p-value <0.05). Stransky et al (2006) do not describe all the genes in the sliding window but only the significant genes in the 28 regions of LRES. Figure 3.5B shows that using the scores generated with my TCS algorithm the genes described in Stransky et al (2006) as having correlated expression with their neighbouring genes also show significantly high scores in the data I generated compared with the rest of the genes on the array (Wilcox $p < 2.2e^{-16}$) .

However, I also found a number of "extra" regions that were not described in Stransky et al (2006). It is possible that additional criteria were used to exclude these regions which are not described in the publication. One contributing factor could be that in their analysis they first identified copy-number dependent regions of coordinate expression and then from these regions eliminated those containing a copy number aberrations, whereas I used the z score approach (rather than permutations) on data that was copy number independent to begin with and therefore had more regions from the start. Also Stransky et al (2006) describe only genes with significant scores and not all the genes in each region. Given that it is seldom possible to reproduce analysis on published microarray data 100% of the time (Ioannidis et al., 2010) I was confident that as I have recapitulated a significant proportion of their data (Fisher $p<2.2e^{-16}$) and produced TCS maps akin to those in their supplementary materials online, I had successfully verified the method for use in my project.

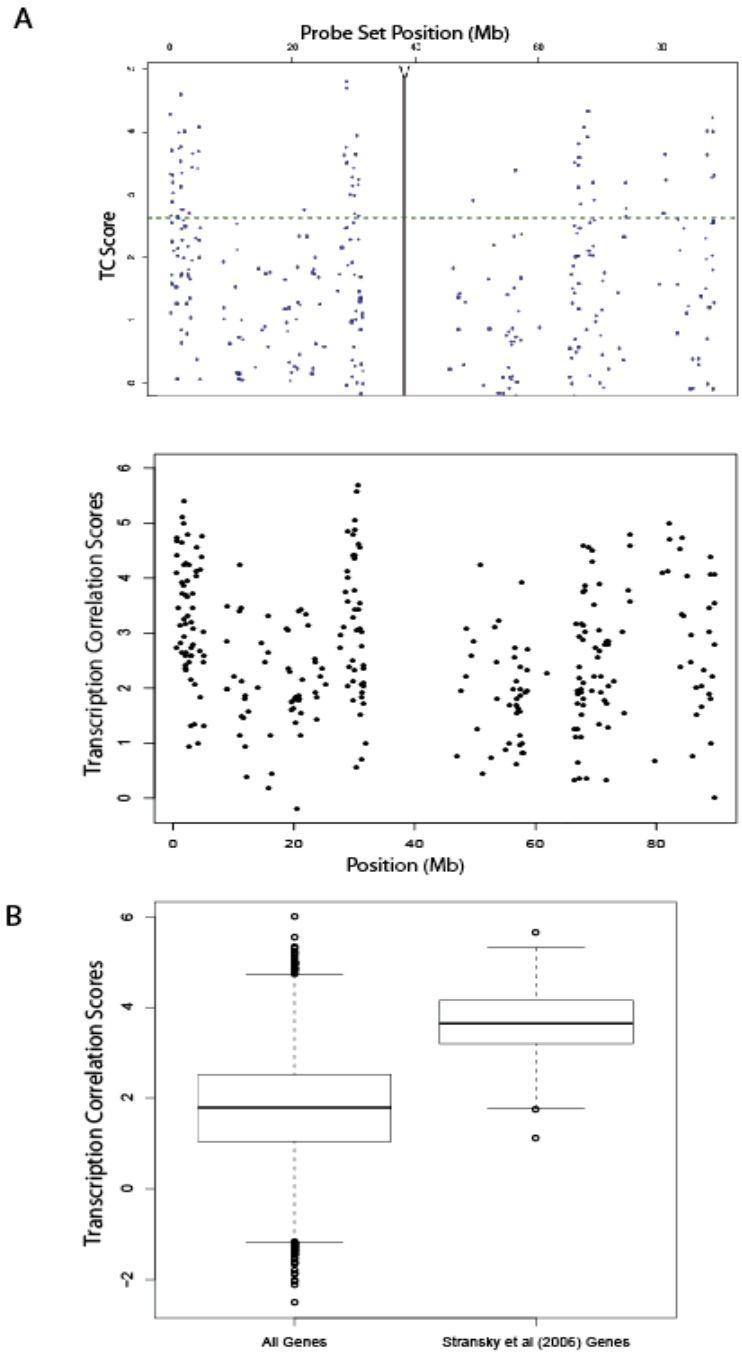


Figure 3.5: Published data on bladder carcinoma is reproducible using an algorithm for identifying regions of coordinate expression. (A) Transcription Correlation Map (TCM) for chromosome 16 produced by Stransky et al (2006) (upper panel) matches the TCM pattern I produced for this chromosome (bottom panel). (B) Genes from Stransky et al., (2006) show significantly higher Transcription Correlation (TC) scores in the data I generated compared to the other genes on the array ($p<2.2e-16$)

3.3.4: The Effect of Varying the Window Size

The number of genes with significant scores ($p<0.05$) was calculated for $2n$ = 1 to 22 neighbouring genes (Figure 3.6). As the number of neighbours increases beyond n=10 the number of significant genes I detected plateaued. Therefore I chose a window size of n=10 for all of my breast tumour analysis. Having selected n=10 for my sliding window algorithm I wanted to confirm that the regions derived at this window size were true regions of transcriptional correlation and therefore assessed whether they appeared in analyses when varying the window size. To do this I made a heat map of all the genes on the array and marked genes with significant correlation scores whilst varying the number of genes in the sliding window (Figure 3.7, black). At smaller window sizes, for example n=1, there are lots of significant TCS genes where 3 genes are co-ordinately being expressed which are not observed at larger window sizes. By comparison, at n=10 the regions are much more consistent (wider horizontal black blocks) where significant TCS genes are co-ordinately expressed at multiple window sizes.

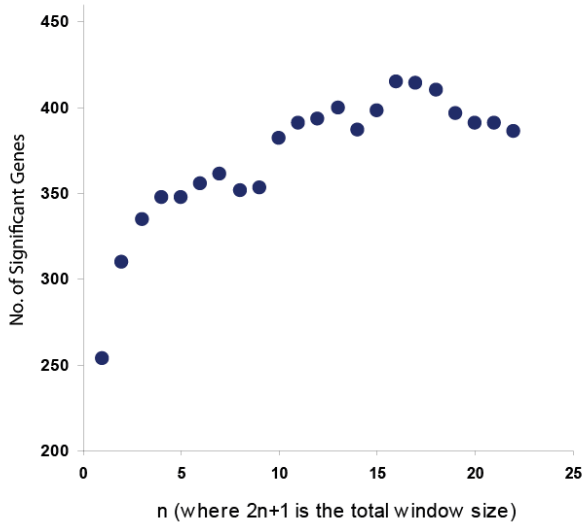


Figure 3.6: The Number of Significant Genes with varying Window sizes. Data for 356 breast tumours showed that varying the window size varies the number of significant genes in the analysis. The number of genes with a significant TCS plateaux's after n=10.

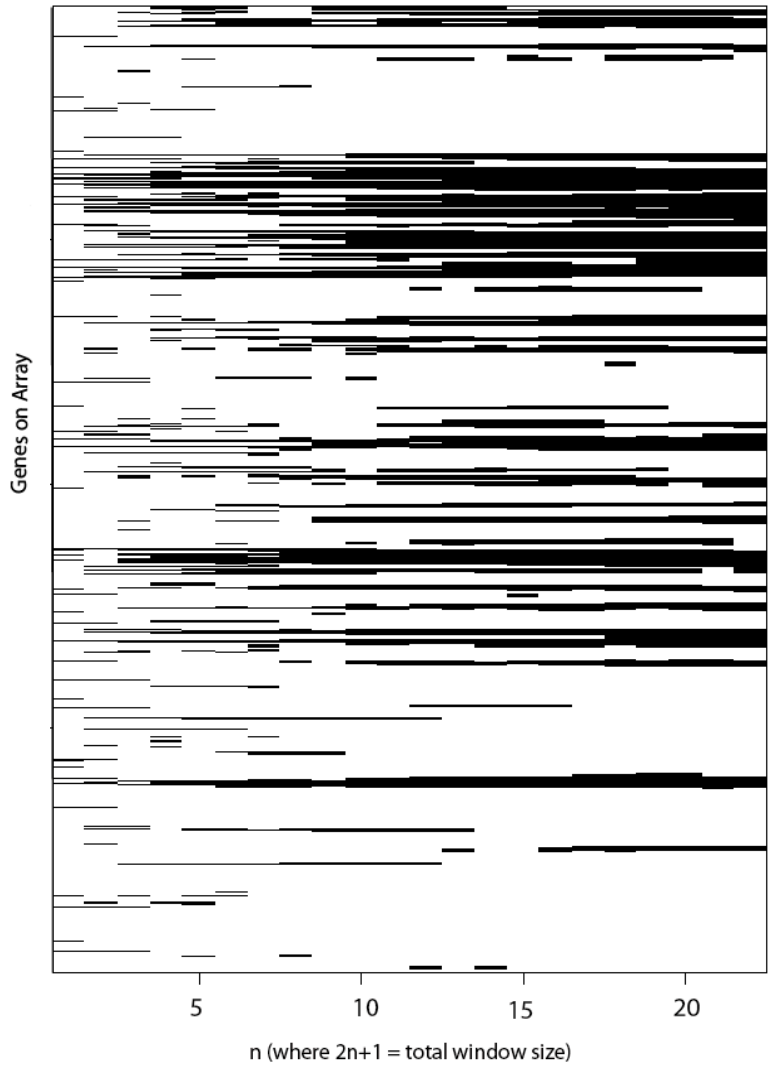


Figure 3.7: Genes with significant TCS (black) when varying the number of neighbours. Plotted is a heat map representation of all genes tested for transcriptional correlation (y axis) at all window sizes (x axis). Genes with significant transcription correlation scores are highlighted black (horizontal bars). Many of the genes are significant across multiple window sizes (black bars). At *n*=10 many of the genes that have a significant TCS at other window sizes are also significant in this gene window.

As I was investigating the effects of high order chromatin structure on gene expression I was expecting the regions to consist of large co-ordinately regulated clusters of genes. Given that the Radvanyi group investigated the same phenomena (Stransky et al., 2006) I examined the size of the regions described in their paper against those that I derived (Figure 3.8). They describe using *n*=7 in the sliding window approach, i.e. 14 neighbours. Therefore initially I did the same, comparing the region sizes for breast tumour data at *n*=7 with the published data (Figure 3.8A). The range of sizes appeared similar. I then

examined *n*=10 for the breast tumour data and the Reyal et al 2005 tumour dataset (Figure 3.8B). This showed a significant difference at *n*=10 between these breast tumour datasets. A number of contributing factors could have caused this but I expect it is because the regions I analysed were copy-number independent whereas Reyal et al (2005) regions will have included any large gain or losses of genetic material that led to coordinate gene expression signatures.

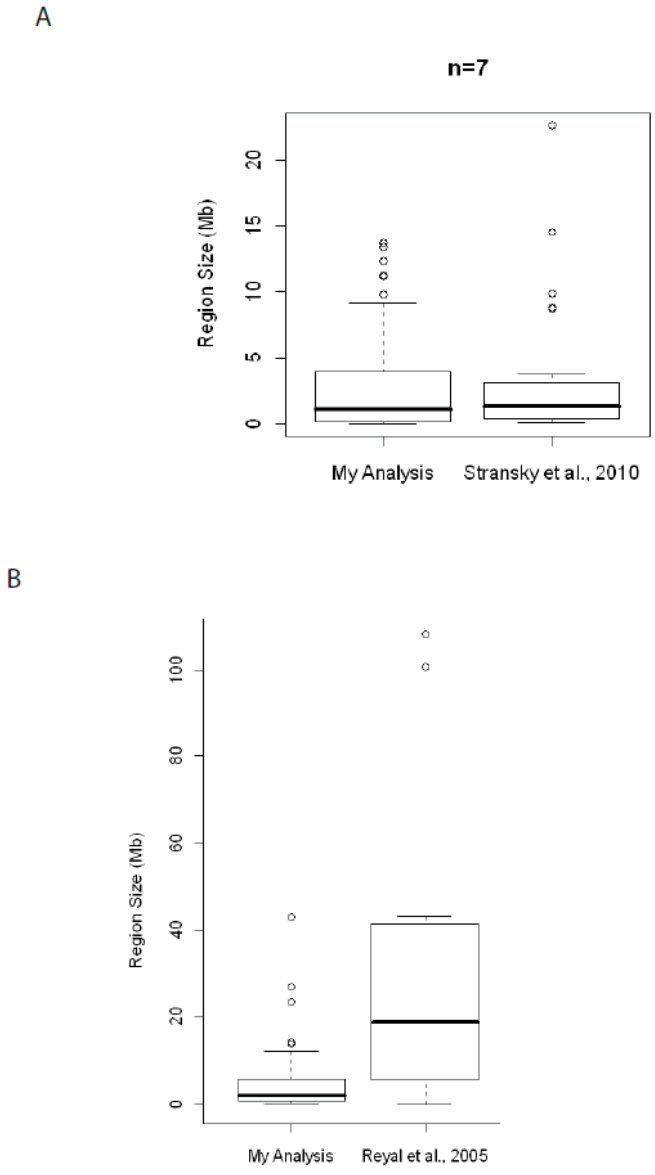


Figure 3.8: Comparison of Regions Sizes to published Coordinately Expressed Regions. (A) Analysis of Bladder Carcinoma Data (Stransky et al., 2006) using a total window size of 15 genes (*n*=7) yields similar size range of RER regions to those published for this window size (B) Analysis of breast tumours (Jonsson et al., 2010) at a total window size of 21 (*n*=10) yields much smaller RER region sizes than published coordinately expressed regions in breast cancer (Reyal et al., 2005)

3.3.5: RER in Breast Tumours

I then used the approach described above to identify regions which showed coordinate changes in gene expression independent of copy number changes in 356 breast tumours (RERs) (Jönsson et al., 2010). Maps of the TCS data for chromosomes 3 and 16 are shown in Figure 3.9 as examples (see Appendix III for all chromosomes). Overall I identified 406 genes with significantly correlated expression to their neighbours ($p<0.05$, threshold score 5.081).

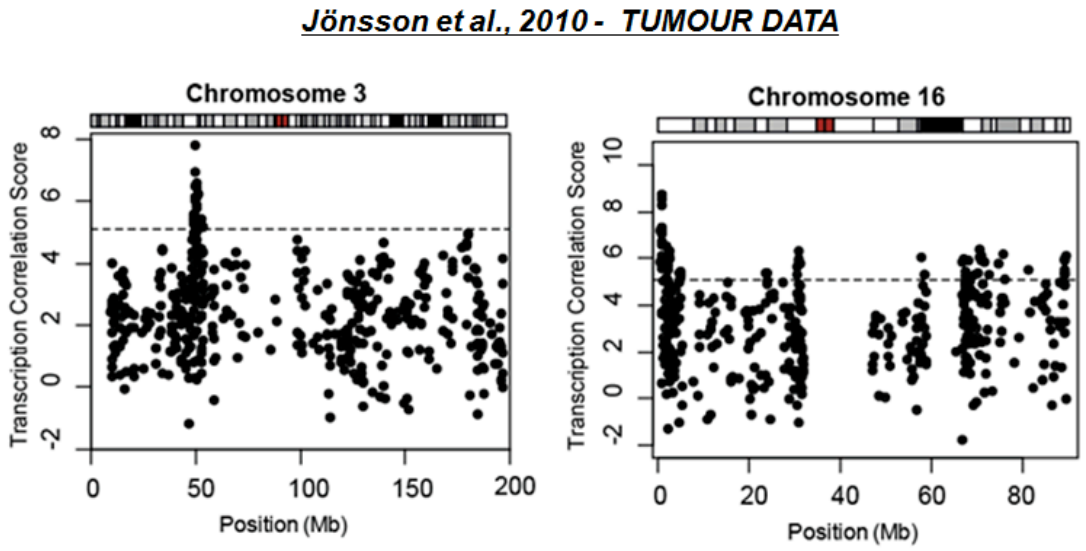


Figure 3.9: Breast tumour Transcription Correlation Maps (TCM's). Transcriptome correlation maps for chromosomes 3 & 16 generated by plotting all TCS values against chromosome position in megabases (Mb), using data for 356 tumours (Jönsson et al., 2010). The dotted line represents the significance threshold. Genes above this line are considered to have expression profiles that are considerably correlated with their neighbours.

This number was reduced to 382 significant genes after removing regions which contain only 1 correlated gene. After merging overlapping regions this finally led to the identification of 45 regions of potential long-range regulation (Table 3.1). The sizes of these regions varied from 0.12 Mb to 43 Mb (median 1.86 Mb). My algorithm for generating transcriptional correlation scores does not take into the distance between genes (and therefore the presence of centromeres or gene desserts). Indeed three of the regions I identified (6, 33 and 39) span centromeres.

Table 3.1: RER domains generated from breast tumour data. The cytogenetic band where the RER's are located, the RER size (measured as the first to last significant TCS in the region) and also the gene ID's for all the significant TCS gene's in the region are shown.

Region	Cytogenetic band	Size (Mb)	No. of Sig. TCS	Significant TCS Genes in Region
1	2p25.1 - 2p24.2	2.35	2	*ADAM17, PDIA6*
2	2q14.2 - 2q22.1	2.91	4	*IWS1, SAP130, UGGT1, IMP4*
3	3p21.31 - 3p14.3	4.45	26	*ZNF589, CCDC51, UQCRC1, IP6K2, P4HTM, WDR6, DALRD3, IMPDH2, QARS, USP19, LAMB2, TCTA, DAG1, APEH, RNF123, IP6K1, RBM5, HYAL2, TUSC4, CYB561D2, TMEM115, MAPKAPK3, TEX264, ABHD14A, BAP1, SPCS1*
4	4q28.2 - 4q31.22	0.29	2	*SCOC, ELMOD2*
5	6p22.3 - 6p22.1	1.39	3	*C6orf62, GMNN, HIST1H4C*
6	6p11.2 - 6q21	27	14	*SLC17A5, TMEM30A, ZNF292, RARS2, MDN1, CASP8AP2, MAP3K7, KIAA0776, NDUFAF4,C6orf167, FBXL4, USP45, CCNC, ASCC3*
7	6q23.2 - 6q25.3	14.27	12	*IFNGR1, HEBP2, C6orf115, HECA, VTA1, PEX3, FUCA2, LTV1, SHPRH, PPIL4, RMND1, C6orf211*
8	7q11.23 - 7q21.2	12.23	4	*STYXL1, MDH2, DMTF1, SRI*
9	7q33 - 7q36.2	10.94	11	*SLC37A3,NDUFB2, MRPS33, SSBP1, CASP2, CUL1, EZH2, ZNF746, ABCB8, TMUB1, CHPF2*
10	8p21.2 - 8p12	0.54	2	*ZNF395, INTS9*
11	8p11.23 - 8q11.23	3.56	6	*TM2D2, GOLGA7, MYST3, AP3M2, IKBKB, C8orf40*
12	8q21.13 - 8q24.3	42.91	32	*NBN, OTUD6B, RAD54B, KIAA1429, ESRP1, INTS8, PLEKHF2, MTERFD1, PTDSS1, MTDH, HRSP12, AP003355.2, VPS13B, ANKRD46, UBR5, AZIN1, ATP6V1C1, SLC25A32, TTC35, TAF2, MRPL13, DERL1, ATAD2, WDYHV1, TRMT12, RNF139*
13	8q24.3	0.99	13	*TSTA3, SCRIB, PUF60, NRBP2, PARP10, GRINA, GPAA1, CYC1, SHARPIN, FBXL6, GPR172A, VPS28, CYHR1*
14	9p21.3 - 9p13.3	1.17	2	*NOL6, SIGMAR1*
15	9q34.11 - 9q34.12	0.78	4	*TRUB2, ODF2, WDR34, DOLPP1*
16	10p15.1 - 10p12.1	5.79	5	*UPF2, CDC123, HSPA14, RPP38, STAM*
17	10q25.3 - 10q26.2	3.11	3	*SEC23IP, PLEKHA1, IKZF5*
18	11q12.2 - 11q13.1	1.709	5	*C11orf48, WDR74, OTUB1, NUDT22, RPS6KA4*
19	11q13.1 - 11q13.2	0.577	6	*KAT5, FIBP, CCDC85B, SART1, SF3B2, YIF1A*
20	12q15 - 12q21.33	8.18	2	*RAB21, PPP1R12A*
21	13q14.11 - 13q14.2	1.32	2	*ESD, MED4*
22	14q23.3 - 14q32.11	4.81	3	*COMMD6, UCHL3, SPRY2*
23	13q33.1 - 13q34	0.42	2	*ANKRD10, ARHGEF7*
24	14q11.2 - 14q12	2.83	7	*CHD8, RBM23, PCK2, DCAF11, RNF31, IPO4, CHMP4A*
25	14q23.3 - 14q32.11	8.36	8	*SLC39A9, COX16, NUMB, ZNF410, COQ6, ENTPD5, GSTZ1, SNW1*

26	16p13.3	1.86	23	*TMEM8A, NME4, RAB11FIP3, PIGQ, RAB40C, LA16c-398G5.2, WDR90, RHOT2, WDR24, METRN, FAM173A, CCDC78, NARFL, IFT140, NME3, MRPS34, HAGH, NDUFB10, GFER, NTHL1, TRAF7, MLST8, E4F1*
27	16p13.3 - 16p13.13	0.45	4	*DNAJA3, ANKS3, ROGDI, UBN1*
28	16p12.3 - 16p12.1	0.12	3	*EARS2, NDUFAB1, PALB2*
29	16p11.2	0.53	6	*TBC1D10B, ZNF48, PRR14, FBRS, PHKG2, BCL7C*
30	16q12.2 - 16q24.1	23.58	22	*COQ9, C16orf57, C16orf80, CMTM1, DYNC1LI2, CES2, TMEM208, ACD, CENPT, SLC7A6OS, CIRH1A, WWP2, AARS, DDX19A, COG4, SF3B3, FTSJD1, DHX38, PSMD7, TMEM170A, ADAT1, CENPN*
31	16q24.1 - 16q24.3	0.51	6	*CYBA, MVD, CTU2, FAM38A, CDT1, ACSF3*
32	17p13.3 - 17p13.2	0.16	2	*MYBBP1A, PELP1*
33	17p12 - 17q11.2	10.47	17	*COPS3, MED9, SMCR8, AKAP10, IFT20, POLDIP2, UNC119, PIGS, SPAG5, KIAA0100, SDF2, SUPT6H, TLCD1, ERAL1, FLOT2, DHRS13, NUFIP2*
34	17q12 - 17q21.2	2.44	17	*TADA2A, MRPL45, MLLT6, PIP5K2B, CWC25, FBXL20, MED1, CRKRS, STARD3, PGAP3, ERBB2, C17orf37, GRB7, GSDMB, ORMDL3, PSMD3, MED24*
35	17q21.2 - 17q21.31	0.79	7	*GHDC, COASY, FAM134C, VPS25, CCDC56, PSME3, AARSD1*
36	17q21.32 - 17q24.1	13.73	33	*PDK2, PPP1R9B, MRPL27, LRRC59, RSAD1, TOB1, MSI2, MRPS23, CUEDC1, SFRS1, DYNLL2, RNF43, MTMR4, RAD51C, TRIM37, C17orf71, DHX40, CLTC, PTRH2, TMEM49, TUBD1, HEATR6, USP32, APPBP2, BCAS3, BRIP1, MED13, DCAF7, STRADA, CCDC47, DDX42, FTSJ3, PSMC5*
37	17q25.1 - 17q25.3	2.72	24	*TMEM104, ICT1, KCTD2, ARMC7, NT5C, HN1, NUP85, MIF4GD, SLC25A19, GRB2, KIAA0195, SAP30BP, GALK1, WBP2, TRIM47, TRIM65, FBF1, C17orf106, EVPL, SRP68, PRPSAP1, UBE2O, RHBDF2, SEPT9*
38	17q25.3	1	13	*AZI1, FSCN2, C17orf70, NPLOC4, HGS, MRPL12, ASPSCR1, STRA13, LRRC45, DCXR, RFNG, GPS1, CCDC57*
39	18p11.22 - 18q11.2	1.06	5	*PSMG2, PTPN2, CEP192, C18orf19, RNMT*
40	18q12.3 - 18q21.32	7.14	3	*IER3IP1, SMAD4, POLI*
41	20q13.2 - 20q13.33	3.82	5	*RAB22A, VAPB, GNAS, TH1L, LSM14B*
42	22q11.23 - 22q12.1	1.18	5	*C22orf13, SNRPD3, C22orf36, KIAA1671, ADRBK2*
43	22q12.2 - 22q12.3	3.03	3	*C22orf28, FBXO7, MCM5*
44	22q13.1 - 22q13.2	0.31	2	*ZC3H7B, PPPDE2*
45	22q13.31 - 22q13.33	0.32	2	*CRELD2, TRABD*

To validate my results I also calculated copy-number-independent TCS's for a second independent breast tumour dataset (Chin et al., 2006) and confirmed that the genes with significant TCS's in one tumour dataset (Jönsson et al., 2010) also produced significantly high TCS's in the second tumour dataset ($p<0.05$). Figure 3.10 depicts this analysis as boxplot distributions of all TCS's generated from each dataset and then the distribution of TCS's for RER genes (as derived from Jönsson et al., 2010 data) in both datasets.

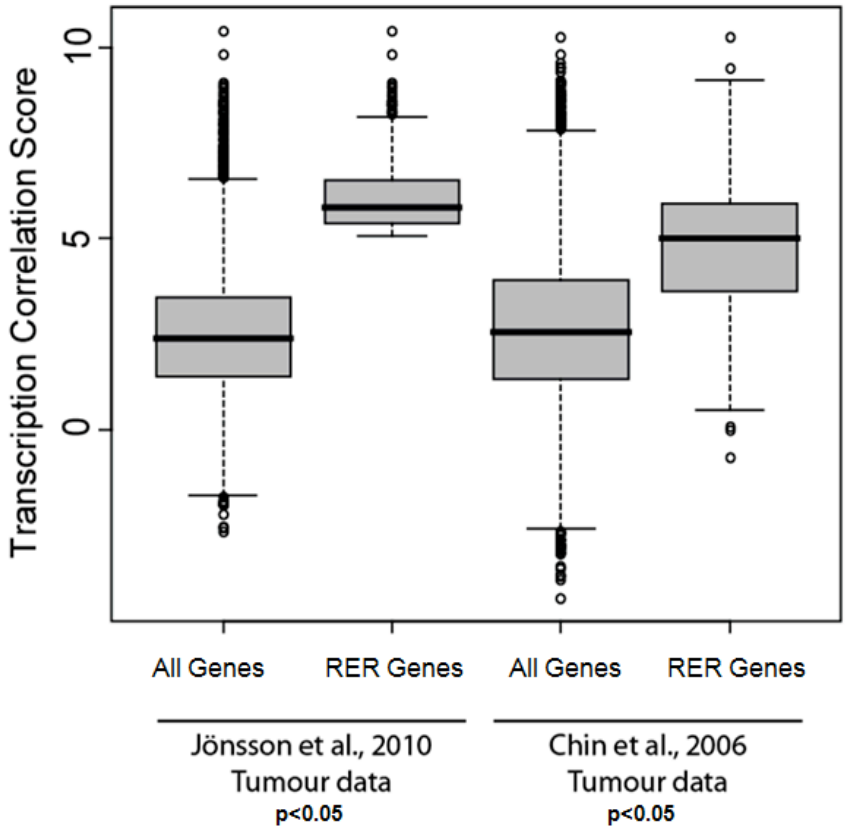


Figure 3.10: Validation of tumour RER genes in a independent datasets. Significant genes generated from Jönsson et al., 2010 tumour data also show significantly high scores in independent tumour tissue data from Chin et al. 2006 (all $p<0.05$).

Having established that these RER domains were true regions of co-ordinately expressed genes I then set out to establish how they were behaving relative to normal breast tissue. Although the dataset I used to derive the 45 RER domains contained a vast number of tumour samples (356) with both expression and CGH data available (Jönsson et al., 2010) there were no normal samples in this dataset. Therefore I used another published dataset containing expression data for 42 breast tumours and 143 normal breast tissues (Chen et al., 2009). In order to compare expression in the 45 RER domains I generated mean-centered z-scores for all genes and used the median z score per sample to plot the distribution as boxplots. The Wilcox test was used to determine whether or not there was a significant difference between tumour and normal samples. I found that not all my regions conformed to the LRES phenotype dominant in published studies. In fact of the 45 regions I identified, 28 were aberrantly upregulated (shaded red in Table 3.1) in tumours compared to normal breast tissue, 6 were aberrantly down-regulated (shaded green in Table 3.1) and 11 showed no significant change in expression. Figure 3.11 shows two examples of RERs with activated gene expression in tumours on chromosomes 3p21.31 and 16p11.2 (top panel) and two repressed regions on chromosomes 13q14 and 13q33 (bottom panel). Appendix IV shows all 45 RER domains.

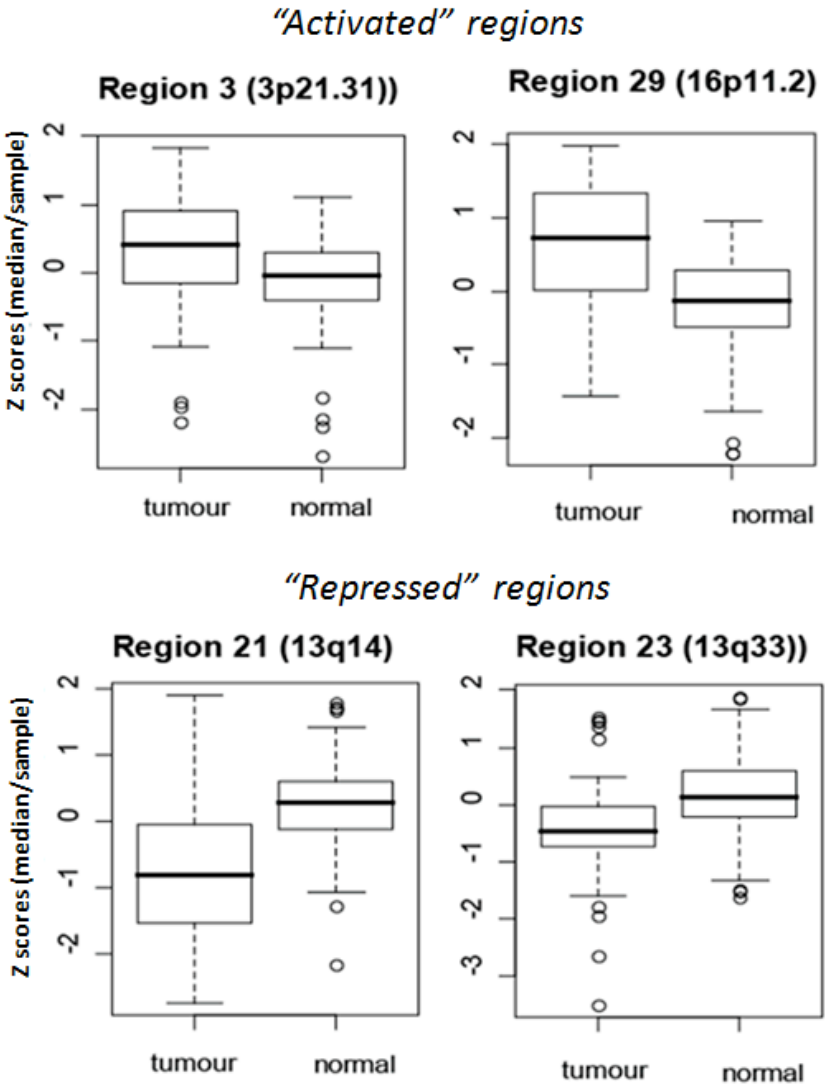


Figure 3.11: Analysis of gene expression changes in tumours relative to normal tissue. Mean centred z scores were calculated for all genes in the region. The median z score for each sample was used to plot the distribution of these scores as boxplots and to compare scores between samples. Regions 3 (3p21.31) and 29 (16p11.2) (top panel) are two examples of regions were genes are co-ordinately upregulated in tumours relative to normal. Regions 21 (13q14) and 23 (13q33), are two examples of regions with coordinate downregulation in tumours.

3.3.6: RER in Breast Cancer Cell Lines

Although analysis of RER domains in tumours would be the 'gold standard', I wanted to move onto mechanistic studies which are not feasible in tumours, to determine how these RER domains are regulated. Therefore having identified 45 RER domains in breast tumours I next set out to identify them in breast cancer cell lines. Using the same approach as used for breast tumours, I analyzed expression and copy number data for breast cancer cell lines from Neve et al. (2006) (Figure 3.12).

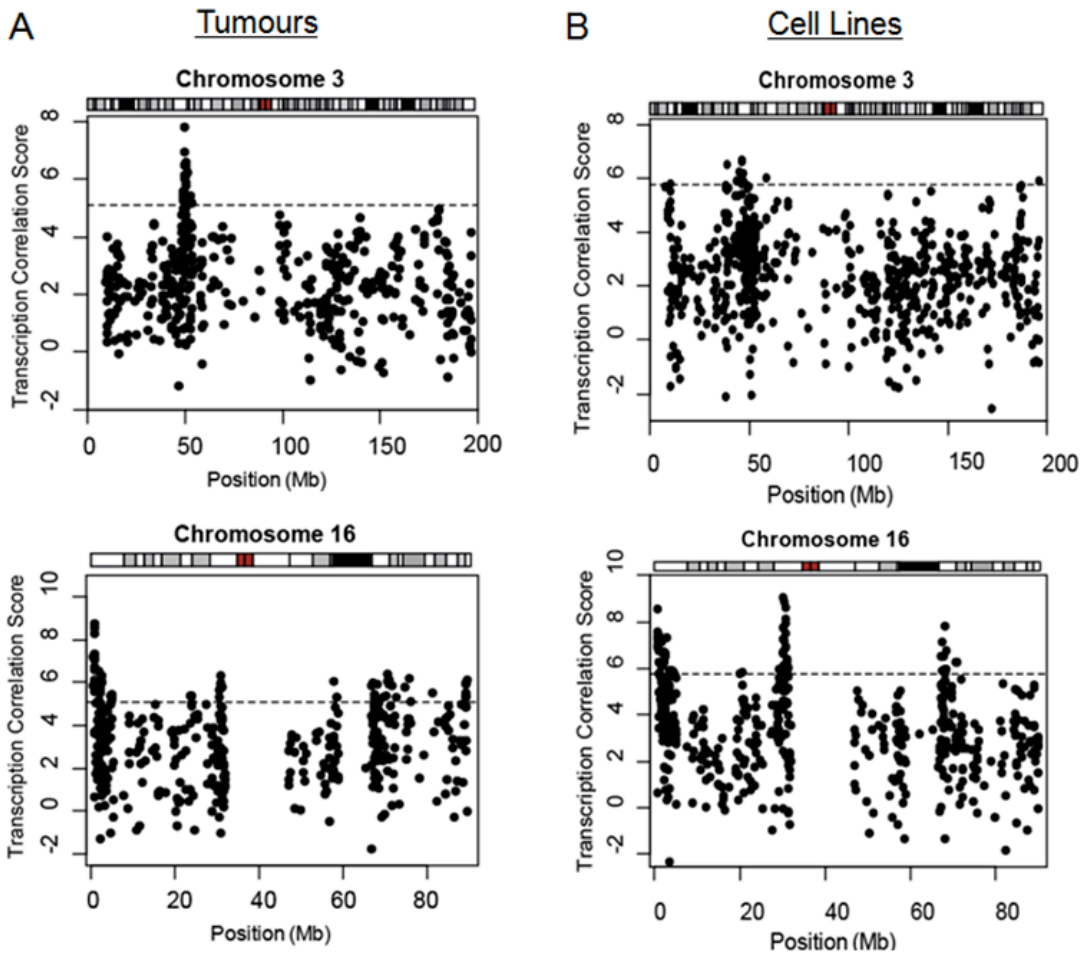


Figure 3.12: Comparison of Transcription Correlation Maps (TCM's) for Breast Cancer Tumours and cell lines. Transcriptome Correlation Maps generated for chromosomes 3 & 16 using data generated for (A) 356 breast tumours (Jönsson et al., 2010) and (B) for 48 breast cancer cell lines (Neve et al., 2006). Dashed line represents the significance threshold.

There were 557 genes that showed significant correlation scores ($p<0.05$, threshold score 5.779) using the Spearman rank-sliding window approach. Figure 3.12 shows examples (chromosomes 3 & 16) of maps of the transcriptional correlation scores in both the tumour and cell line datasets. Both the breast tumour and cell line maps show similar peaks of gene clusters with significantly high TCSs. This gives me confidence that breast cancer cell lines can be used to investigate the molecular mechanisms underlying RERs. Appendix V shows the TCS maps for all chromosomes derived from this cell line analysis.

After excluding those regions containing only 1 correlated gene, 501 RERs remained in the cell line data. Merging of overlapping regions produced 71 regions with sizes ranging from 0.01 Mb to 15.8 Mb (median 0.9 Mb) (Table 3.2).

Table 3.2: RER domains generated from breast cancer cell line data. The cytogenetic band where the RER's are located, the RER size (measured as the first to last significant TCS in the region) and also the gene ID's for all the significant TCS gene's in the region are shown. Blue shading indicates RER domains that overlapped with those derived from the tumour analysis.

Region	Cytoband	Size (Mb)	No. of Sig. TCS	Significant TCS Genes in Cell Line Region
1	1p36.13	0.532719	2	*AKR7A3, RNF186*
2	1p34.3	0.332406	3	*CDCA8,YRDC,UTP11L*
3	1p34.2 - 1p34.1	0.627772	3	*ELOVL1,ATP6V0B,B4GALT2*
4	1q21.3 - 1q23.3	7.545616	42	*SPRR1B,S100A14,SLC39A1,JTB,C1orf43,UBAP2L,HAX1,UBE2Q 1,ADAR,PMVK,ADAM15,EFNA3,EFNA1,RAG1AP1,DPM3,MUC1, SCAMP3,CLK2,RUSC1,YY1AP1,DAP3,ROBLD3,RAB25,MEF2D, GPATCH4,BCAN,CD1C,CD1B,CD1E,CADM3,CTA34P22.2,DARC ,OR10J1,APCS,CRP,SLAMF8,CCDC19,KCNJ10,KCNJ9,ATP1A2, CASQ1,CD84*
5	1q42.2	0.37173	2	*TTC13, GNPAT*
6	1q43	0.392338	2	*FH,EXO1*
7	2q32.3 - 2q33.1	9.658744	4	*MYO1B, HSPE1, C2orf47, NIF3L1*
8	3p22.2	0.330516	2	*VILL, SLC22A14*
9	3p22.1 - 3p21.31	3.457713	9	*ANO10, ZDHHC3, SLC6A20, FYCO1, CCR1, CCR2, CCR5, CCRL2, PRSS50*
10	4p16.3	2.047026	11	*TACC3, WHSC1, WHSC2, HAUS3, RNF4, TNIP2, SH3BP2, ADD1, TETRAN, HTT, ADRA2C*
11	4q13.2 - 4q13.3	5.110343	11	*UGT2A3, SULT1B1, CSN1S1,CSN2, STATH, CSN3, SMR3B, PROL1, IL8, CXCL5, CXCL3*
12	4q22.1	0.398405	2	*MEPE, ABCG2*
13	4q26 - 4q27	2.091682	3	*MYOZ2,PDE5A,TNIP3*

14	5p13.2	2.465914	4	*RAD1,BRIX1,SKP2,NUP155*
15	6p22.2 - 6p22.1	3.613025	34	*HIST1H4B,HIST1H3B,HIST1H2AB,HIST1H2BB,HIST1H3C, HIST1H1C,HIST1H2BC,HIST1H2AC,HIST1H1E,HIST1H2BD ,HIST1H2BE,HIST1H4D,HIST1H3D,HIST1H2BF,HIST1H4E, HIST1H2BG,HIST1H2AE,HIST1H1D,HIST1H3F,HIST1H2BH, HIST1H2APS4,HIST1H2BI,HIST1H4H,HIST1H2BM,HIST1H2AM ,HIST1H2BO,ZNF435,GPX5,OR2W1,OR2J3,OR2N1P,OR12D2, OR10C1,MOG*
16	6p21.33	0.072086	4	*LST1,NCR3,BAT2,APOM*
17	6p21.33 - 6p21.32	0.962297		*TNXB, CREBL1, PPT2, GPSM3, NOTCH4, C6orf10, HLA-DRB6, HLA-DQB2, HLA-DMA*
18	6q23.2	0.146092	3	*TAAR5,TAAR3,VNN3*
19	6q25.2	1.348873	3	*RP3-468K3.1, RP3-527B10.1,OPRM1*
20	6q26 - 6q27	6.429087	7	*PARK2,PDE10A,T,CCR6,GPR31,TCP10,C6orf123*
21	7p22.3 - 7p22.1	3.965987	6	*MAD1L1,FTSJ2,CHST12,FOXK1,ACTB,RNF216*
22	7q22.1	0.146208	2	*EPO,SLC12A9*
23	8p21.3	0.904261	6	*BMP1,POLR3D,PIWIL2,SLC39A14,PPP3CC,SORBS3,BIN3, RHOBTB2,TNFRSF10B*
24	8p11.21 - 8q11.23	13.05061	6	*AP3M2,POLB,VDAC3,SLC20A2,MCM4,MRPL15*
25	8q13.3 - 8q21.3	15.859747	6	*KCNB2,STAU2,FAM164A,STMN2,FAM82B,MMP16*
26	8q22.2 - 8q22.3	5.291133	11	*RPL30,HRSP12,NIPAL2,VPS13B,COX6C,SPAG1,RNF19A, ANKRD46,UBR5,AZIN1,FZD6*
27	9p13.3	0.723551	6	*KIAA1045,DNAJB5,RUSC2,CD72,SIT1,CA9*
28	9q34.3	2.096118	5	*OLFM1,C9orf116,SNAPC4,ABCA2,GRIN1*
29	10p13	1.979364	3	*HSPA14,NMT2,RSU1*
30	10q22.1 - 10q22.2	1.995287	4	*PSAP,CBARA1,SEC24C,NDST2*
31	10q23.33	0.386544	2	*CYP2C18, CYP2C8*
32	10q26.11 - 10q26.13	2.780985	5	*BAG3,C10orf119,SEC23IP,BRWD2,PLEKHA1*
33	11q12.1 - 11q12.2	0.931709	6	*MS4A2,MS4A5,MS4A12,GPR44,TMEM109,CD6*
34	11q13.1	0.191506	2	*COX8A,MACROD1*
35	11q13.2	0.158078	2	*SPTBN2,C11orf80*
36	11q23.3 - 11q24.3	11.105129	14	*FXYD6,CD3G,MLL,H2AFX,PDZD3,THY1,TECTA, SCN3B,ACRV1,DDX25,CDON,KCNJ1,KCNJ5,TP53AIP1*
37	12p13.32	0.157215	2	*AKAP3,GALNT8*
38	12p13.31	0.952419	3	*C3AR1,AICDA,KLRG1*
39	12p13.2	1.458118	2	*TAS2R9,LRP6*
40	12q13.11 - 12q13.12	0.718107	3	*ASB8,CACNB3,RND1*
41	12q13.2	1.293331	4	*PDE1B,BLOC1S1,RDH5,MMP19*
42	12q21.31 - 12q22	7.055034	4	*NTS,MGAT4C,DCN,EEA1*
43	14q32.13	0.228806	3	*SERPINA2,SERPINA4,SERPINA5*
44	16p13.3	1.948303	23	*AXIN1,TMEM8A,NME4,RAB11FIP3,PIGQ,RAB40C ,RHOT2,FAM173A,NARFL,GNG13,LMF1,CACNA1H, TPSG1,TPSD1,UBE2I,BAIAP3,NME3,MRPS34,NUBP2, GFER,NTHL1,PGP,E4F1*
45	16p13.3	1.068754	5	*OR1F1,OR2C1,NAT15,ADCY9,TFAP4*

46	16p12.3	0.621226	2	*C16orf88,GP2*
47	16p11.2	1.224821	25	*SEZ6L2,TAOK2,HIRIP3,DOC2A,ALDOA,PPP4C,TBX6, MAPK3,CD2BP2,TBC1D10B,ZNF771,ZNF768,ZNF747, ZNF764,ZNF688,ZNF785,PRR14,FBRS,SRCAP,PHKG2, RNF40,ZNF629,CTF1,SETD1A,VKORC1*
48	16q22.1 - 16q22.2	3.893046	11	*CDH16,NOL3,E2F4,ATP6V0D1,THAP11,PSKH1,DDX28 ,DUS2L,PRMT7,COG4,VAC14*
49	17p11.2	0.39805	3	*TOM1L2,LRRC48,LLGL1*
50	17q11.2	0.155523	4	*UNC119,KIAA0100,SDF2,SUPT6H*
51	17q21.2	0.363708	6	*KRTAP1-3,KRTAP1-1,KRTAP2-4,KRTAP4-9,KRT34 ,KRT31*
52	17q21.31	0.659717	2	*RUNDC3A,C1QL1*
53	17q25.1	0.358511	4	*KCTD2,GGA3,MRPS7,GRB2*
54	17q25.3	0.315342	4	*STRA13,RFNG,CSNK1D,SECTM1*
55	19p13.3	0.662667	4	*PIAS4,ZBTB7A,SH3GL1,C19orf10*
56	19p13.2	0.030115	2	*MAP2K7,SNAPC2*
57	19p13.2 - 19p13.12	5.058953	23	*CDC37,TMED1,AC024575.1,EPOR,ELAVL3,CNN1, ACP5,C19orf56,TNPO2,ASNA1,BEST2,PRDX2,RNASEH2A ,MAST1,DNASE2,GCDH,GADD45GIP1,PRKACA,ASF1B ,OR7C2,SLC1A6,EPHX3,WIZ*
58	19p13.11	1.329896	8	*SLC5A5,PIK3R2,PGPEP1,ELL,FKBP8,DDX49,TMEM161A ,RFXANK*
59	19q13.12	0.55062	6	*CD22,GAPDHS,ZBTB32,ARHGAP33,NPHS1,APLP1*
60	19q13.2	0.717892	7	*CD79A,GRIK5,ZNF574,POU2F2,GSK3A,CIC,CEACAM8*
61	19q13.33 - 19q13.42	4.526636	37	*SHANK1,KLK1,KLK3,KLK2,KLK6,KLK11,KLK12,KLK13, KLK14,SIGLEC9,SIGLEC7,CD33,NKG7,LIM2,SIGLEC6, ZNF175,SIGLEC5,HAS1,FPR1,FPR2,FPR3,ZNF528,LILRB3, LILRB5,LILRA3,LILRA5,LAIR1,KIR3DX1,LILRB1,LILRB4, AC006293.3,KIR2DL4,KIR3DL1,FCAR,NCR1,GP6,SYT5*
62	19q13.43	0.461506	5	*ZNF550,ZNF134,ZNF211,ZNF586,ZNF606*
63	20p13	0.013104	2	*OXT, AVP*
64	20p11.23 - 20p11.1	7.381602	14	*RBBP9,INSM1,FOXA2,CYB5P4,THBD,CD93,CST8,CST3, CST4,CST5,TMEM90B,CST7,C20orf3,FAM182B*
65	20q13.12	0.483177	4	*ZSWIM1,MMP9,CDH22,SLC35C2*
66	20q13.33	0.820961	3	*DIDO1,ARFGAP1,RTEL1*
67	21q22.3	0.669345	2	*RRP1, LRRC3*
68	22q11.21 - 22q12.1	6.914663	13	*TRMT2A,P2RX6,TOP3B,PPIL2,IGLV1-40,ZNF280B,ZNF280A, ZDHHC8P,VPREB3,MMP11,UPB1,SEZ6L,CRYBB1*
69	22q12.2	0.169807	2	*INPP5J, PIK3IP1*
70	22q12.3 - 22q13.1	0.751707	4	*TMPRSS6,SSTR3,MFNG,GCAT*
71	22q13.1	1.548667	6	*CBX6,APOBEC3A,PDGFB,MGAT3,CACNA1I,SGSM3*

3.3.7. Overlap of RER's between cell line and tumours

I found that many of the RERs identified in breast tumours overlapped completely or in part with RERs from the breast cancer cell line datasets (Figure 3.12; Table 3.1; Table 3.2 blue shading; Figure 3.13). There were also instances, for example RER 33 in the tumour data (which spans the centromere on chromosome 17), where the region could be divided into two smaller regions based on the two regions from the cell line data (regions 49 & 50) which it overlapped with either side of the centromere. In addition, a region on chromosome 16p11.2 which appeared in both the tumour and cell line RER analyses encompasses the LRES domains previously identified in breast cancer cells (Hsu et al 2010).

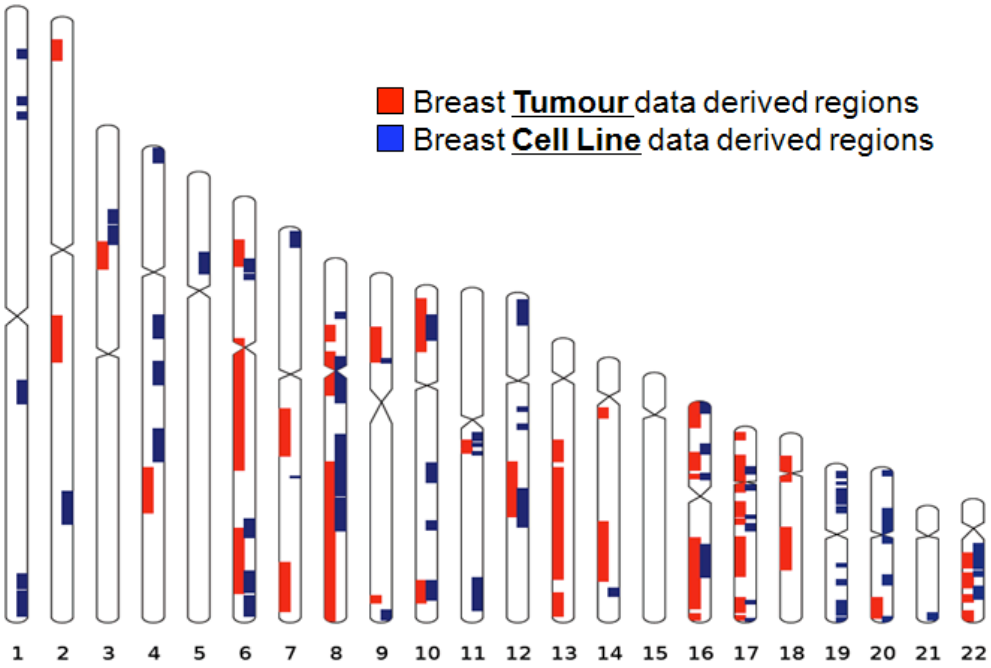


Figure 3.13: Comparison between RERs derived from tumour and cell line data. Ideogram representation of chromosomal locations of 71 RER's in breast cancer cell lines (blue) and 45 RER's in tumours (red).

In total I was able to identify 26 copy number independent regions of coordinate expression that are in common between breast tumours and breast cancer cell lines based on their overlap (Figure 3.13; Table 3.1). The overall size range of these 26 regions was 0.23-13.4 Mb (mean 3.67 Mb , median 1.40 Mb). Appendix VI contains tables of all genes within the 26 RER coordinates. Gene Ontology (GO) analysis for genes within the

overlapping regions showed a high enrichment a number of interesting GO terms including nucleosome assembly (GO:0006334), nucleosome organization (GO:0034728) and chromatin organization (GO:0006325). A full list of associated GO terms are included in Appendix IX.

Table 3.3: 26 RER domains derived from overlap between tumour and cell line RER's. The cytogenetic band where the RER's are located, the RER size (measured as the first to last significant TCS in the region), the total number of genes within each region according to Ensembl (including the entire gene window and those not included in the arrays), whether or not the regions are misregulated in tumours and the p-value associated with this analysis.

Region	Cytoband	Size (Mb)	All Genes within RER coordinates (Ensembl)	Misregulation of Gene Expression (tumour v. normal)	Wilcox.test (p-val) for tumour v. normal
1	6p22.2	0.89	66	UP	0.000000
2	6q23.2 - 6q23.3	3.43	60	DOWN	0.003904
3	6q25.1 - 6q25.3	6.71	86	DOWN	0.000023
4	8p11.21 - 8q11.23	13.41	86	DOWN	0.000023
5	8q21.13 - 8q21.3	10.54	118	No Sig. Change	0.8428
6	8q22.1 - 8q23.1	11.99	212	No Sig. Change	0.1806
7	9p13.3	1.13	51	No Sig. Change	0.1108
8	10p13 - 10p12.31	6.71	105	DOWN	0.04647
9	10q26.11 - 10q26.13	6.34	99	No Sig. Change	0.09559
10	11q12.3 - 11q13.1	0.75	39	UP	0.000038
11	11q13.2	0.36	28	UP	0.000015
12	12q21.31 - 12q21.33	10.39	94	DOWN	0.007449
13	16p13.3	2.16	182	UP	0.000000
14	16p13.3	1.45	57	UP	0.000000
15	16p12.3	0.38	10	No Sig. Change	0.8996
16	**16p11.2**	**1.36**	**124**	**UP**	**0.000000**
17	16q22.1 - 16q22.2	4.9	225	DOWN	0.03697
18	17p11.2	1.69	58	DOWN	0.009758
19	17q11.2	0.53	37	UP	0.000008
20	17q21.2	0.28	12	No Sig. Change	0.4965
21	17q25.1	0.9	41	UP	0.000006
22	17q25.3	0.58	37	UP	0.000000
23	22q11.23 - 22q12.1	5.93	164	DOWN	0.04232
24	22q12.2 - 22q12.3	1.01	43	No Sig. Change	0.4402
25	22q12.3	0.24	6	No Sig. Change	0.2172
26	22q13.1 - 22q13.2	1.37	42	No Sig. Change	0.4559

Taking these overlapping regions between tumour and cell lines as domains that are potential candidates for further mechanistic study I sought to determine how they were behaving in tumour vs normal tissue using data from (Chen et al., 2009). From this analysis I found once again that the regions did not conform to the LRES phenotype that is dominant in published studies.

Of the 26 regions I found that, for nine of them gene expression was upregulated relative to normal breast and was downregulated in a further eight regions. There were also nine regions that showed no significant change in gene expression between tumour and normal tissues. Examples of aberrantly up/down-regulated regions are shown in Figure 3.14 and highlighted red/green respectively in Table 3.3 with Wilcox p-values shown for all 26 regions. Boxplots for all 26 regions are shown in Appendix VII.

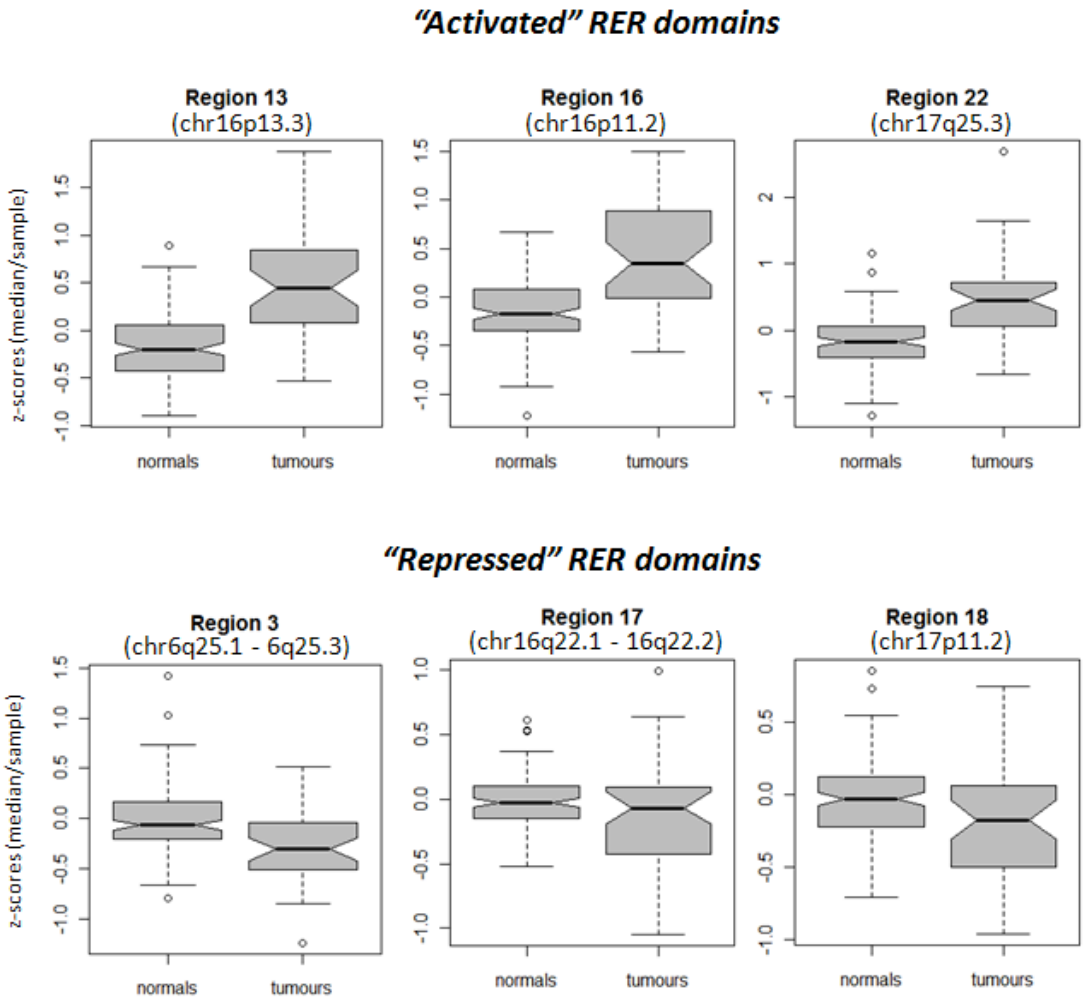


Figure 3.14: Analysis of gene expression changes in tumours relative to normal tissue. Mean centred z scores were calculated for all genes in the region. The median z score for each sample was used to to plot the distribution of these scores as boxplots and enabling relative comparison of scores between samples. The top panel shown, regions 3 and 29, are two examples of regions that were relatively upregulated in tumours. The bottom panel shown, regions 21 and 23, are two examples of regions that were relatively downregulated in tumours.

3.3.8. RER domains and tumour sub-type

Breast cancers fall into a distinct sub-types dependent on their gene expression profiles and the expression of specific receptors (see Section 1.1.2.5 of Chapter 1). To determine if different RERs were characteristic of different tumour sub-types I performed unsupervised hierarchical clustering of gene expression data for all RER regions in both the breast cancer cell line and tumour data. This analysis showed that whilst there is a lot of heterogeneity with regards to the gene expression signature for a region there are also cases where there is a subtype-specific gene signature and in many cases this was sufficient to separate out the main luminal and basal-like subgroups in both the cell lines and tumour data sets. Where this signature exists in the region there is a relative up/down-regulation of the genes which is dependent on the molecular subtype. The data was transformed into mean centred z scores to enable relative comparison of gene expression changes between samples. Red/green indicates an increase/decrease in gene expression relative to the universal mean for each gene.

Figure 3.15 shows an example for RER 13 on chromosome 16p13.3 where in both tumours and cell lines this region is relatively upregulated in the luminal (ER+ve) subtype compared to the basal (ER-ve) subtype tumours where it is relatively repressed. The heat map for the tumour tissue data (Figure 3.15A) showed a much more heterogeneous pattern in gene expression compared with the cell line data. This is likely to be a reflection of the cellularity of the tumour samples and generally the cell line data showed a much better segregation into luminal (ER+ve) Vs basal (ER-ve) gene signatures at RER domains than did the tumour data. Heat maps representing the clustering of gene expression data for all RER domains identified for breast tumours and cell lines in this chapter are available in Appendix VIII.

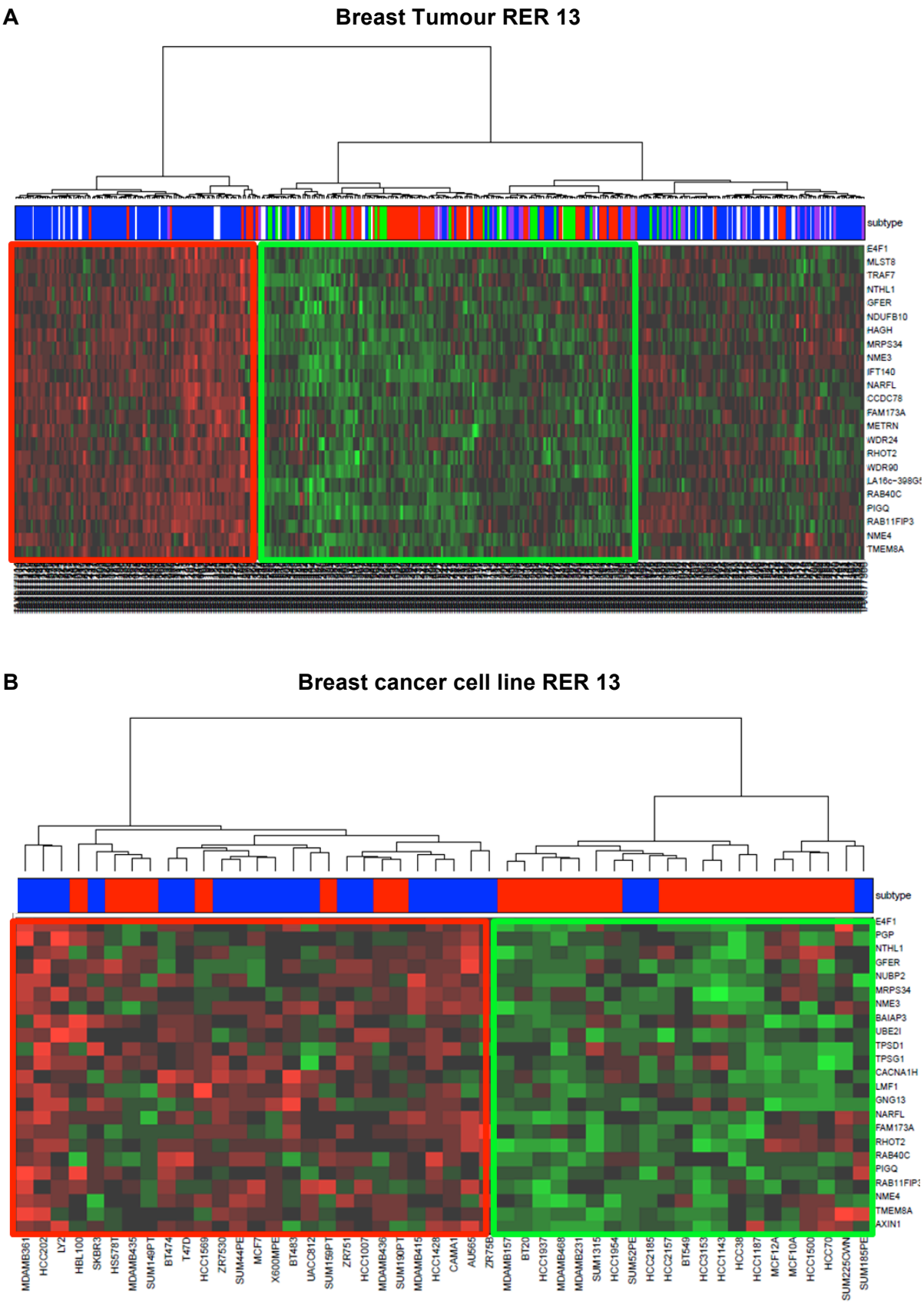


Figure 3.15: Unsupervised cluster analysis of genes in RER 13 on chromosome 16p13.3. Heat maps for genes in RER number 13 to indicate the similarity of sample z scores for breast

tumour samples (A) and cell lines (B). Each tumour or cell line is identified by a colour-coded matrix below the dendrogram representing the results of hierarchical clustering: basal (red), luminal (blue), ERBB2 (purple), normal-like (green). Red/green indicate an increase/decrease in gene expression relative to the universal mean for each gene. The values were mean centred and the colours scaled from the max (+) and min (-) standard deviations for each heat map.

3.3.9: Other properties of RERs

Given that the RERs were co-ordinately regulated and that the sizes of the regions highly variable, I investigated the gene density of the RER domains. This showed that RERs were occurring in regions of the genome that were significantly more gene dense than expected (Figure 3.16A, $p=2.2e^{-16}$). This analysis was carried out initially for the data from the breast tumours, though data from the breast cancer cell lines and bladder carcinomas also showed significant enrichment in regions of high gene density (data not shown).

As a result, I investigated whether the regions of coordinate expression that I observed in breast cancer corresponded to Regions of Increased Density of Gene Expression (RIDGE's) previously identified in the human genome from the examination of gene expression data across 12 normal and pathological tissue types (Caron et al., 2001). RIDGEs are gene dense, have a high GC content and short introns, and anti-RIDGE's have also been identified with opposite properties (Versteeg et al., 2003). Using genome coordinates for RIDGEs and antiRIDGEs obtained from the Versteeg lab I investigated whether or not the tumour tissue RERs correlated with these domains. Indeed many of my RER regions fell into RIDGE's (48%) and also anti-RIDGE's (15%), however not all RIDGEs/anti-RIDGES were regions of epigenetic regulation (RER) (Figure 3.16B). By generating a random set of regions ("fake" RERs) I was able to show that there was no co-occurrance of RER domains with RIDGE's (Fishers p-value 0.395) or antiRIDGE's (Fishers p-value 1) above that expected at random (see Appendix X).

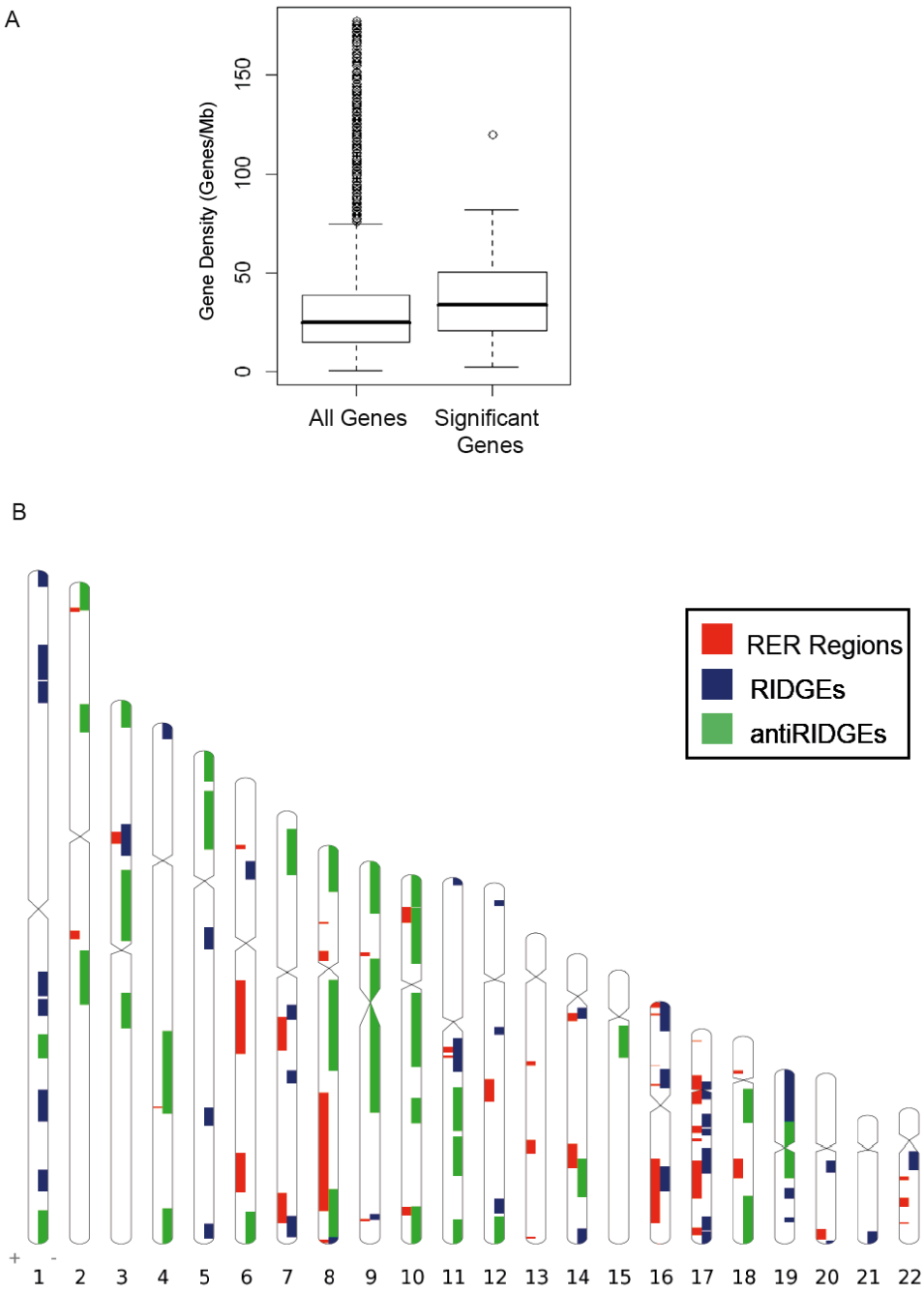


Figure 3.16: Properties of RER regions. (A) Genes with a significant TCS occur in gene dense regions of the genome (calculated as genes/Mb) ($p<0.05$) (B) Ideogram representation of chromosomal locations for tumour RER's (red), RIDGE's (blue) and anti-RIDGEs (green) as defined by Versteeg et al (2003).

3.4: Discussion

In this chapter I developed an approach based principally on that used to identify LRES regions in bladder carcinoma (Stransky et al., 2006). My approach differed in that, whereas Stransky et al (2006) identified regions of coordinate repression and then eliminated regions where coordinate repression was due to copy number, I first removed all copy number affected parts of the dataset before going on to derive regions of coordinate regulation. From this analysis I was able to show that there are at least 26 regions in the breast cancer genome that show coordinate gene expression that this is not due to copy number aberrations. This was shown by analysis of both breast tumour and cell line data and determining which genes had consistently high TCS across both analyses.

Most interestingly the regions identified were not all subject to the LRES phenomena. Indeed a large proportion of these regions that were co-ordinately regulated were aberrantly activated in tumour tissues, and a large proportion showed no difference in their expression relative to normal tissues. At this stage it is also not possible to say that all the RER domains are epigenetically regulated. I have however identified regions which could potentially be regulated by epigenetic mechanisms and could be investigated further.

Selection of a region to further investigate *in vitro* and *ex vivo* was based on the following criteria:

(a) The region would be amongst the 26 overlapping regions between tumour and cell line data analysis
(b) The region would show a degree of subtype specificity in its gene expression signature (as assessed by heat map analysis)
(c) Based on the analysis of tumour and normal samples, the region would show significant aberrant repression or activation
(d) The region selected would be of amenable size to work with
(e) The region when examined on the UCSC browser contained most of the genes in the region.

The RER I identified as fitting these criteria the most was on chromosome 16p11.2. This region as identified in the breast tumour (region 26) and cell line data (region 47) was almost completely overlapping. Figure 3.17 shows a schematic of the RER from the UCSC browser with both the breast tumour and cell line RER marked.

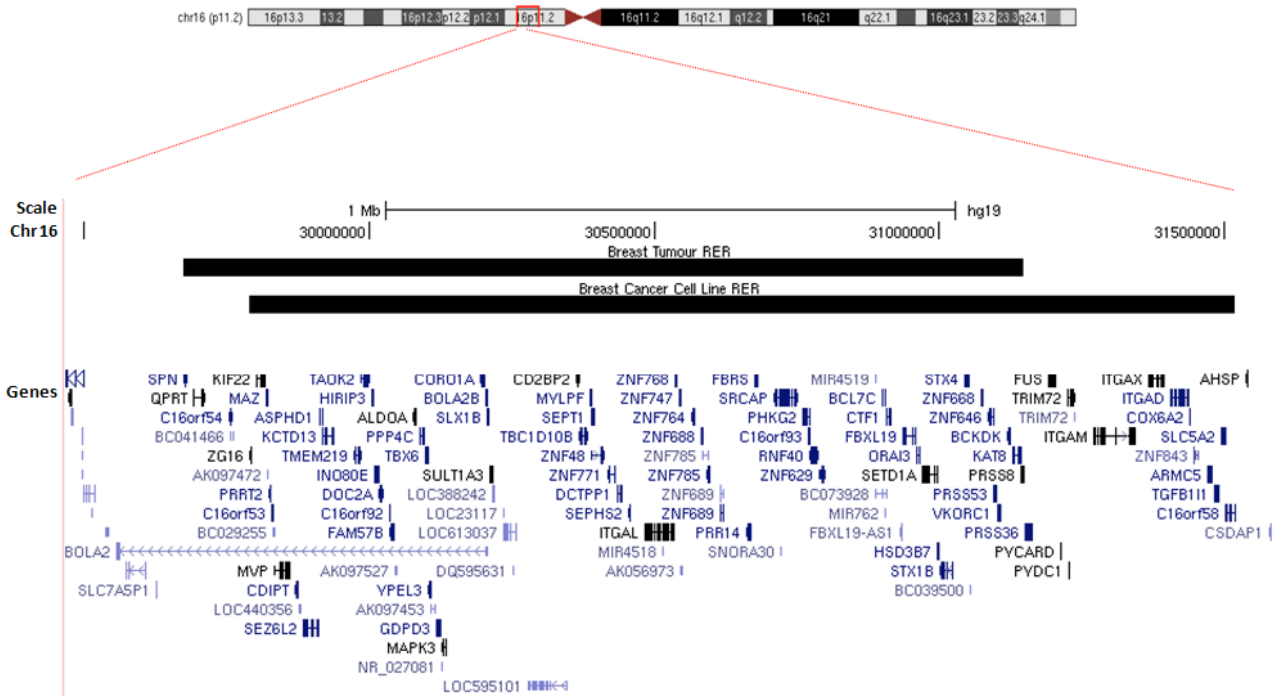


Figure 3.17: Schematic representation of RER on chromosome 16p11.2. Adapted from the UCSC genome browser (build hg19). Black bars represent RER domains in tumours (top) and cell lines (bottom). Ideogram above image indicates location of RER on chromosome 16.

This region is 1.48 Mb in size taking into account all genes in the analysis window and 0.53 Mb taking into account the first to last gene with a significant TCS. This region also contained genes (*QPRT, MAZ, PRRT2, MVP, ASPHD1, TMEM219, TAOK2, INO80E)* previously shown to be involved in estrogen-mediated repression by DNA looping in breast cancer (Hsu et al., 2010). Interestingly this region does not conform to the LRES phenotype described by previous studies. This region appears to be aberrantly upregulated in breast cancer (Figure 3.14: top panel, middle) and shows segregation by heat map analysis of expression data into the two main subgroups of breast cancer. Also, this region contains ZNF proteins and it has been shown that CBX, a component of

heterochromatin that is associated with repression (through binding H3K9me2/3), can coat regions of up to 4Mb that harbour KRAB-ZNF genes (Vogel et al., 2006).

The mechanism by which this region is misregulated in breast cancer is investigated and discussed in chapter 4 and 5 of this thesis.

Chapter 4

An RER on chromosome 16p11.2 shows altered chromatin organisation

4.1: Introduction

Rather than single genes, large regions of coordinately misexpressed genes, termed Long Range Epigenetic Silencing (LRES) regions, have emerged as an exciting new phenomena in the cancer literature and have been attributed to a cocktail of chromatin remodelling and DNA methylation changes (Coolen et al., 2010; Dallosso et al., 2012, 2009; Frigola et al., 2006; Hsu et al., 2010; Javierre et al., 2011; Novak et al., 2008, 2006; Park et al., 2011; Stransky et al., 2006; Vallot et al., 2010). These studies have relied heavily on using expression datasets to identify regions of coordinately misregulated genes and have investigated epigenetic changes at these target loci to identify the possible cause of the misregulation. The LRES status of a tumour has been implicated in the aggressiveness of tumours and therefore also has implications for patient prognoses (Dallosso et al., 2009; Vallot et al., 2010).

In the previous chapter I identified large domains of co-ordinately misregulated genes in breast cancer and established that, as well as LRES, there is another phenomenon occurring which involves long-range activation of expression compared to normal tissues, rather than gene silencing. I therefore denoted domains from my analysis as regions of epigenetic regulation (RER) rather than LRES. Given that there is no consensus from previous studies as to the cause of coordinated misregulation of long stretches of the genome, I hypothesized that changes in chromatin structural organization itself may be causative of gene expression changes for such large regions, as has been seen for some developmentally regulated gene clusters (Eskeland et al., 2010).

In this chapter I describe experiments which set out to investigate chromatin compaction and decompaction as a mechanism of altered gene expression for large misregulated domains (RER's). Additionally I investigated the influence of the nuclear environment by investigating the localisation of a RER with respect to the nuclear periphery, which is known to be the home of silent heterochromatin.

4.2: Results

Fluorescence in situ hybridisation (FISH) is a molecular cytogenetic technique which has been used by the Bickmore lab (and others) to detect whole chromosomes as well specific DNA loci in the nucleus. FISH can be used to measure the compaction of chromatin over a specified genomic region (Section 1.2.4.2)(Sachs et al., 1995). Chromatin compaction has been well studied using this method during differentiation of embryonic stem cells, (Chambeyron and Bickmore, 2004), in the developing embryo (Chambeyron et al., 2005; Morey et al., 2007), comparing wildtype and mutant cells (Ragnhild Eskeland et al., 2010) and also for the study of multiple loci in the same cells (Gilbert et al., 2004). Data from 2D FISH analysis of fixed cells has been shown to recapitulate the chromatin compaction observed with 3D FISH though distances are typically larger (Ragnhild Eskeland et al., 2010; Morey et al., 2007). Therefore in this chapter all cell-line based analyses are based on 2D FISH data which are more swiftly imaged, processed and analysed quantitatively compared with 3D FISH. 3D FISH was used for analyses of tissue sections.

4.2.1: Selection of RER domain on chromosome 16p11.2 to investigate changes in chromatin organisation

As described in chapter 3, a computational approach was used to identify regions of epigenetic regulation (RER) in breast cancer. I identified 45 RER's in breast tumours (Table 3.1) with a view towards studying the mechanism of RER biology and how this relates to transcription. In order to study this phenomenon *in vitro*, I carried out an identical analysis in breast cancer cell lines which revealed 71 RER's (Table 3.2). To ensure that the region I selected to study *in vitro* was a true domain of RER that also occurred in tumours I compared both the tumour and cell line analysis to produce a list of 26 potential target RER's (Table 3.3) that could be studied easily by *in vitro* analysis of cell lines and that are not just cell line artefacts. Comparison of gene expression data for breast tumour and normal tissues revealed 8 LRES regions and 9 aberrantly activated

RERs in tumours. As the published literature focused on LRES regions I decided to focus on up-regulated RER's.

One of these candidate up-regulated regions (RER no.16 in Table 3.3) contained the genes (*SPN*, *QPRT, ZG16, KIF22, MAZ, PRRT2, MVP, SEZ6L2, ASPHD1, KCTD13, TMEM219, TAOK2, INO80E, DOC2A*) that have been reported to undergo ER-mediated looping in breast tumour cells in association with LRES (Hsu et al., 2010). This 1.38 Mb RER domain on chromosome 16p11.2 which overall I found had aberrantly activated gene expression in tumours compared to normal breast (Figure 4.1A) was therefore selected for further study.

This RER was also selected because heat map analysis by z scores of gene expression for all genes in the RER (not just those with a significant transcriptional correlation score) revealed a differential gene expression pattern between the two broad categories of breast cancer subtypes - luminal-type/estrogen-receptor positive (ER+ve) and basal-type/estrogen-receptor negative (ER-ve) - as shown by unsupervised hierarchical cluster analysis (Figure 4.1B).

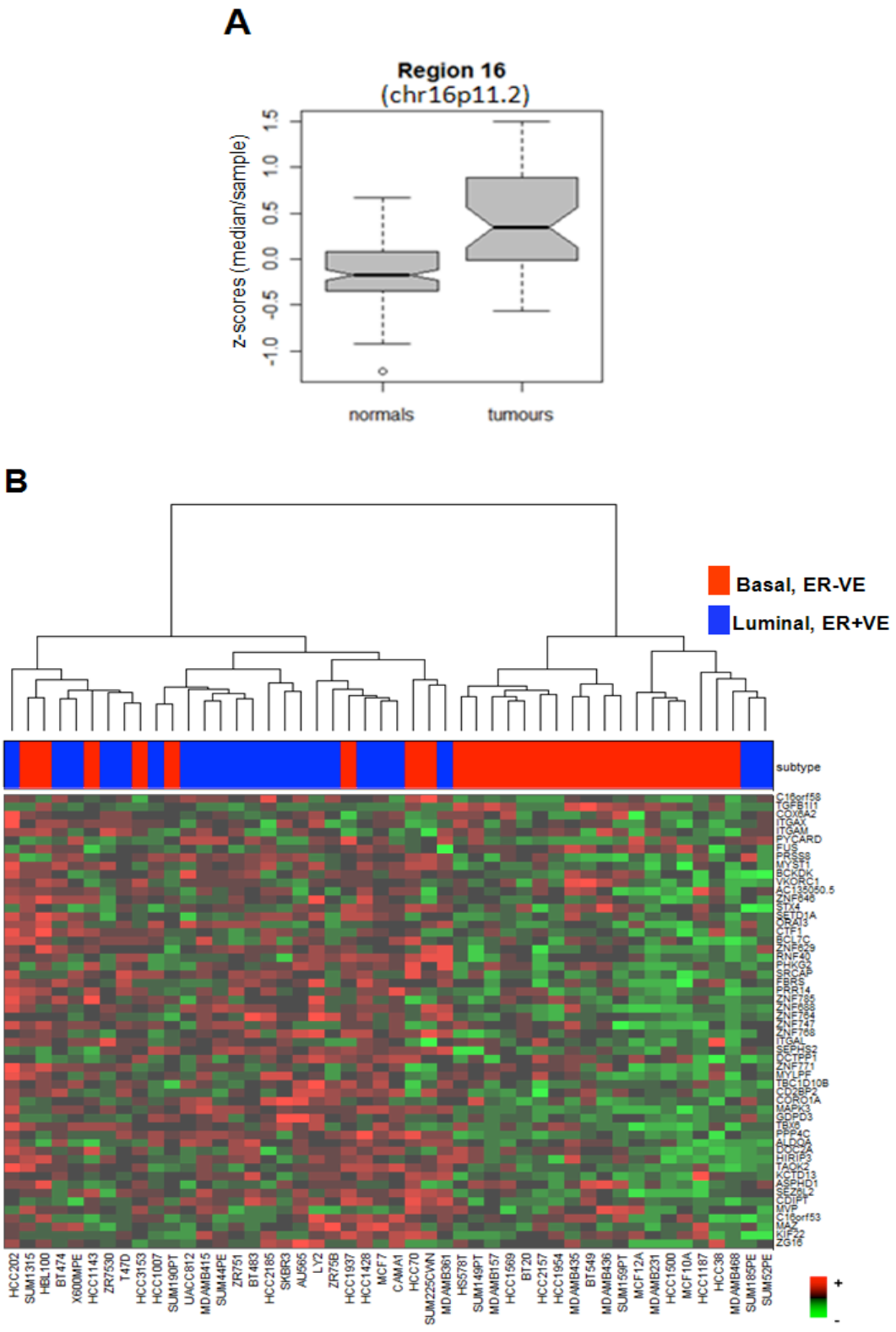


Figure 4.1: Gene Expression at an RER on chromosome 16p11.2 in breast tumours and tumour subtypes. (A) Mean centred z scores were calculated for all genes in the region. Box plot of the distribution of median z scores for all sample compared between normal tissue and tumours. (B) Unsupervised cluster analysis of breast cell lines from Neve et al (2006).The heat map is vertically clustered using Euclidean clustering (shown as dendogram) to indicate the

similarity of sample (z) scores with each other. Based on subtype information from Neve et al each sample is identified by a color-coded matrix below the dendrogram: basal (red), luminal (blue) The legend shows z-scores where red/green indicate an increase/decrease in gene expression relative to the universal mean for each gene. The values were mean centred and the colours scaled from the max (+) & min (-) standard deviations.

4.2.2: Investigating chromatin compaction of RER domain

Chromatin compaction was assayed by measuring the physical distance between pairs of fosmid probes that flanked the locus of interest. Fosmid probes were selected using the fosmid end pairs track of the UCSC genome browser (http://genome.ucsc.edu/). Fosmids flanking each locus were labelled and detected as described in Methods (Chapter 2). The interprobe separations between the "green" and "red" hybridisation signals was determined by in-house scripts. Distances converted from pixels to microns were always squared (d^2) as previous work in the Bickmore lab has shown that in FISH experiments d^2 is proportional genomic distance. Mean d^2 was used to estimate the level of chromatin compaction in all comparisons between cell lines. FISH can be used to measure chromatin compaction at genomic distances of <1Mb, as chromatin follows a random-walk giant-loop mathematical model, with loops above 1.5Mb as discussed in section 4.2.7 (Sachs *et al*, 1995; van den Engh *et al*, 1992).

As the RER domain being investigated is large (1.38 Mb) it was subdivided into four sub-regions of similar size (locus 1: 382 Kb, locus 2: 414.2 Kb, locus 3: 445.2 Kb and locus 4: 418.5 Kb) to accurately assess chromatin compaction across the region (Figure 4.2). Significant differences in chromatin compaction were assessed using the Wilcox test with a cut off of $p<0.05$. This non-parametric test was chosen as it does not rely on the data belonging to a particular distribution but can make a pair-wise assessment between the population mean ranks between datasets.

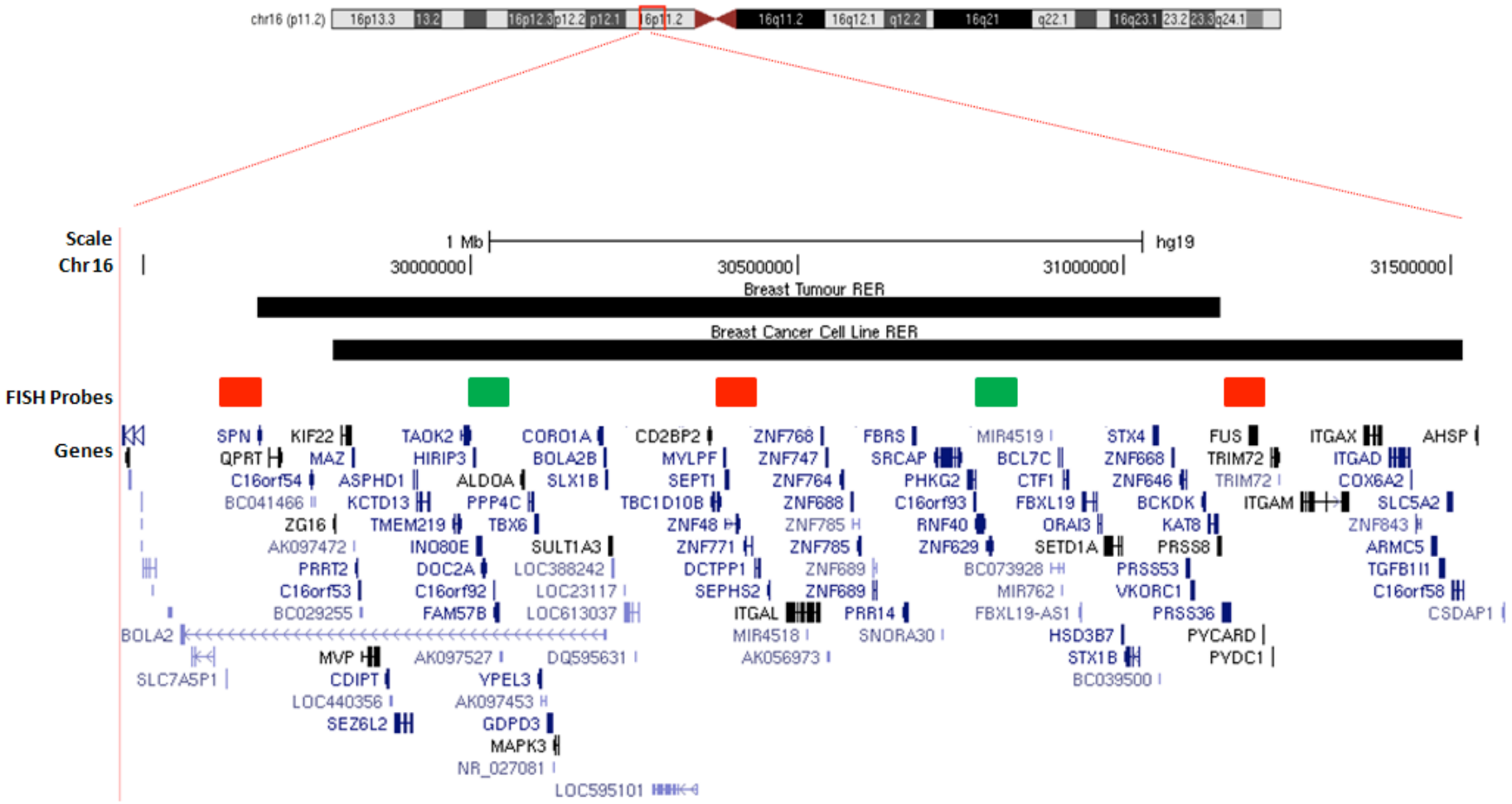

chr16 (p11.2)
16p13.3
13.2
16p12.3
p12.2
p12.1
16p11.2
16q11.2
16q12.1
q12.2
16q21
q22.1
16q23.1
23.2
23.3
q24.1
Scale
Chr16
1 Mb
hg19
30000000
30500000
31000000
31500000
Breast Tumour RER
Breast Cancer Cell Line RER
FISH Probes
Genes
SPN
KIF22
TAOK2
CORO1A
CD2BP2
ZNF768
FBRS
MIR4519
STX4
FUS
ITGAX
AHSP
QPRT
MAZ
HIRIP3
BOLA2B
MYLPF
ZNF747
SRCAP
BCL7C
ZNF668
TRIM72
ITGAD
C16orf54
ASPHD1
ALDOA
SLX1B
SEPT1
ZNF764
PHKG2
CTF1
ZNF646
TRIM72
COX6A2
BC041466
KCTD13
PPP4C
TBC1D10B
ZNF688
C16orf93
FBXL19
BCKDK
ITGAM
SLC5A2
ZG16
TMEM219
TBX6
ZNF48
ZNF785
RNF40
ORAI3
KAT8
ZNF843
AK097472
INO80E
SULT1A3
ZNF771
ZNF785
ZNF629
SETD1A
PRSS8
ARMC5
PRRT2
DOC2A
LOC388242
DCTPP1
ZNF689
BC073928
PRSS53
TGFB1I1
C16orf53
C16orf92
LOC23117
SEPHS2
ZNF689
MIR762
VKORC1
C16orf58
BC029255
FAM57B
LOC613037
ITGAL
PRR14
FBXL19-AS1
PRSS36
CSDAP1
BOLA2
MIR4518
SNORA30
HSD3B7
PVCARD
MVP
AK097527
DQ595631
AK056973
STX1B
PYDC1
SLC7A5P1
CDIPT
VPEL3
BC039500
LOC440356
AK097453
SEZ6L2
GDPD3
MAPK3
NR_027081
LOC595101

Figure 4.2: RER on chromosome 16p11.2 from Breast Tumour and Cell Line Analysis. UCSC human genome browser window showing; Top - ideogram of chromosome 16 with the region (p11.2) containing the RER domain highlighted. Below – expanded view of the RER. The top track shows the extent of the RER identified from the analysis of 356 breast tumours or 48 breast cancer cell lines. The FISH probes browser track (red and green) represents the location of fosmids used as probes for analysis of the region by FISH. At the bottom, the locations of all the genes within the RER is shown.

As the RER I had selected showed good segregation into subtypes by heat map analysis of the gene expression signature (Figure 4.1B) I selected a breast cancer cell line that was luminal ER+ve (MCF7) and another that was basal-type ER-ve (MDAMB231) according to Neve et al (2006) for initial analyses. These two cell lines also fell into separate clusters by unsupervised Euclidean analysis (Figure 4.1B). Unnormalised inter-probe distances (d^2) showed a significant difference (Wilcox test $p<0.05$) for locus 2 and 4 with MCF7 cells being less compact at these loci than in MDAMB231 cells. It appeared that locus 3 also followed this trend though the data was not statistically significant. Locus 1 on the other hand appears to have the same level of chromatin compaction in both cell lines.

The observed inter-probe distances were normalised to the radius squared (d^2/r^2) as previous work as well as my own showed considerable differences in the nuclear area between cell types and preparation of slides which could skew distances when comparing them. When normalised in this way, locus 2 remained consistent in having significantly more decompact chromatin in MCF7 cells compared with MDAMB231 cells (Figure 4.3). Though locus 4 appears from the boxplots to also show this pattern the effect is not significant. Locus 3 has the exact same distribution of normalised inter-probe distances between cell lines. Locus 1 appears to be more decompact in MDAMB231 cells than in MCF7 cells, an affect that would perhaps be expected taking into account the proposed ER-mediated looping of genes in the region to form a compact structure (Hsu et al., 2010), however this effect is not significant in either normalised or unnormalised FISH data.

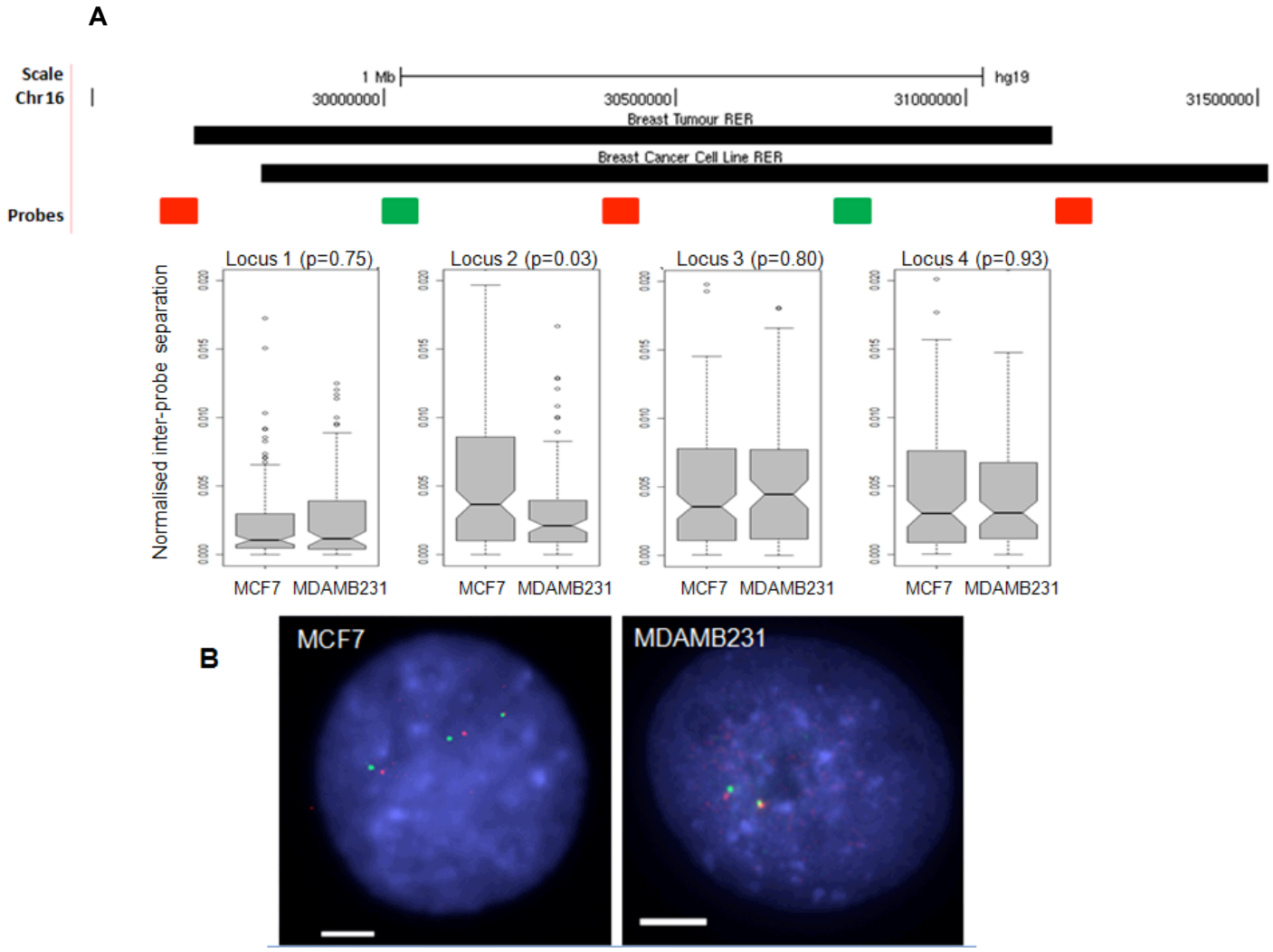


Figure 4.3: FISH analysis of compaction for RER on chromosome 16p11.2. (A) Top - the location of the RER identified from the analysis of 356 breast tumours and 48 breast cancer cell lines with the location of FISH probes. Below - boxplots show the distribution of normalised interprobe distances in MCF7 and MDAMB231 cells. Shaded boxes show the median and interquartile range of data; circles represent outliers. n= 45-60 nuclei. (B) Representative FISH images of MCF7 and MDAMB231 nuclei hybridised with probes W12-1754H9 (green) and W12-906G10 (red) from subregion 2. DNA is stained with DAPI (blue). Scale bars = 5μm.

4.2.3: Further delineation of the RER domain and subtype specific changes in chromatin compaction

In order to further investigate the locus 2 region of the 16p RER, where I detected different chromatin compaction between MCF7 and MDAMB231 cells, I revisited the heat map analyses of gene expression data selecting only for the genes in the locus 2 region. Unsupervised hierarchical clustering of locus 2 using expression data from the 48 breast cell lines (Neve et al., 2006) was sufficient to segregate the z scores into two main branches that corresponded to the two main sub-classes of breast cancer: luminal-type ER+ve and basal-type ER-ve cancer (Figure 4.4). From this heat map it is clear that the two selected cell lines, MCF7 and MDAMB231, were consistently repressed/activated relative to one an other across the entire locus.

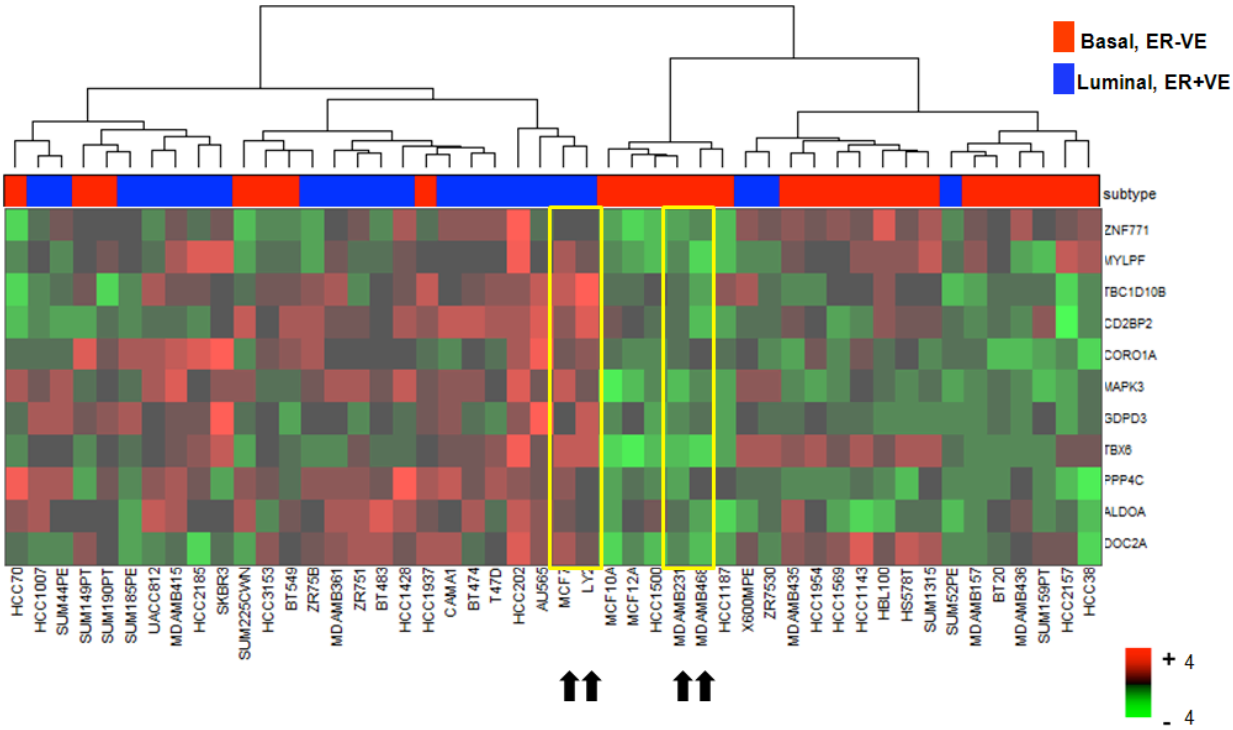


Figure 4.4: Unsupervised cluster analysis of Locus 2 in the 16p11.2 RER. The heat map is vertically clustered using Euclidean clustering (shown as dendogram) to indicate the similarity of sample scores with each other. Based on subtype information from Neve et al., 2006 each sample is identified by a color-coded matrix below the dendrogram: basal (red), luminal (blue) The legend shows z-scores where red/green indicate an increase/decrease in gene expression relative to the universal mean for each gene. The values were mean centred and the colours scaled from the max (+4) & min (-4) standard deviations. This enables relative comparison of scores between samples. Yellow box and arrows indicate cell lines used for FISH.

I therefore returned to the computational analysis described in chapter 3 to determine whether this region on chromosome 16p11.2 could be further delineated using a smaller window size. Varying the number of neighbours in the analysis from *n=1-10* where *2n+1* is the total window size I was able to identify all the possible genes that showed correlated gene expression patterns with the other genes in the window. I then plotted a heat map of all genes with a significant transcription correlation score when decreasing the window size (Figure 4.5). This analysis shows that the high scoring genes cluster into groups across the RER on chromosome 16p11.2, one of which is mostly in locus 2 extending into locus 1 and the other of which lies in locus 3.

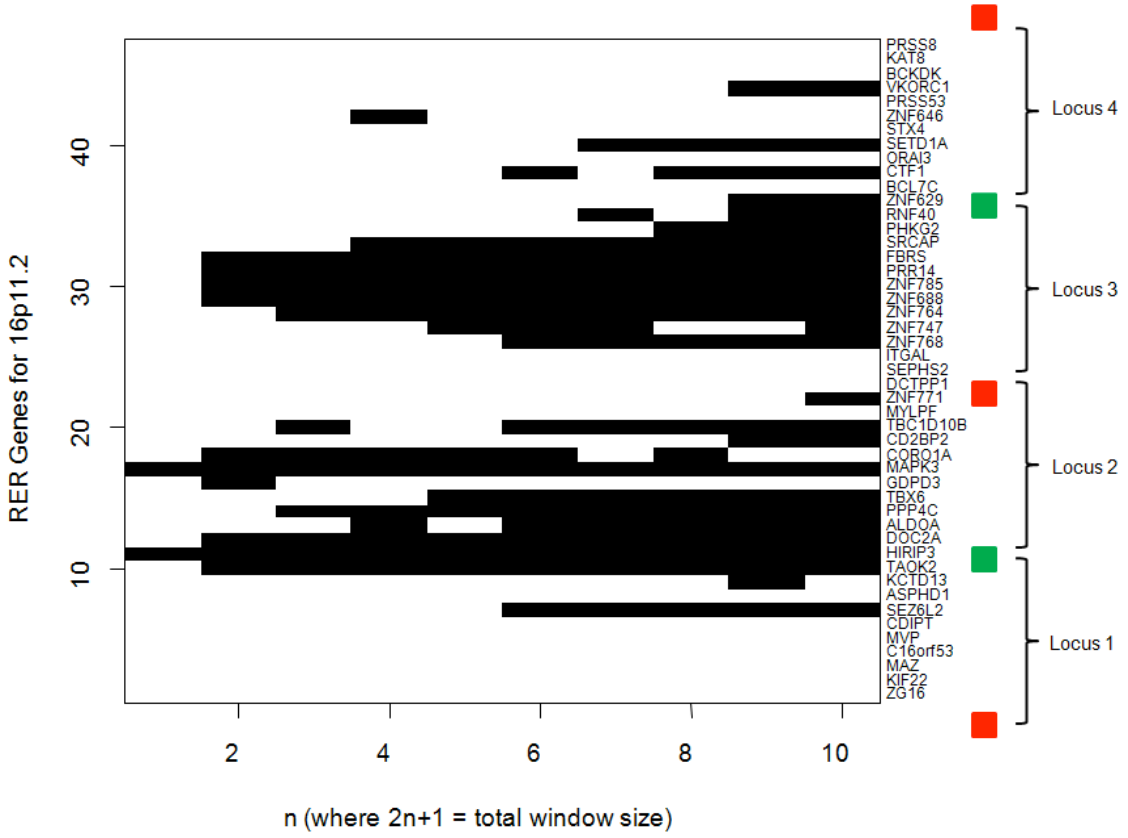


Figure 4.5: Heat map representation of 16p11.2 RER with varying window size. Only genes in the region with significant transcription correlation scores (TCS) are coloured black on the heat map at each of the sliding window sizes (x axis). Genes across the region (y axis) are named on the right of the heat map. Adjacent to gene names are the FISH probes demarking the 4 loci that were examined. This figure is based on the breast cancer cell line derived TCS's and map of genes on the expression arrays used by Neve et al (2006).

To determine the entire region of transcriptional correlation for these two clusters of significant TCS genes, I delineated the regions as described in Chapter 3 – by taking all the genes in the analysis window for all significant TCS genes, discarding regions that only had one significant TCS and therefore producing a list of raw regions for each window size. The resulting regions at 16p11.2 are shown in Figure 4.6 for all window sizes, alongside the location of the refined tumour and cell line derived RERs (as described in chapter 3).

This segment of chromosome 16p11.2 does appear to be a valid RER at different sliding window sizes. Between *n=5-10* this remains as one large RER domain that then splits into two smaller RERs at *n=2-4*. One of these small RER domains (left) covers completely (and infact spans out of) locus 2 whereas the other smaller RER (right) is mostly but not entirely contained in locus 3 (Figure 4.6). It is therefore not unexpected that of the 4 loci examined by FISH it was locus 2 where the difference in chromatin compaction was seen and not the other loci.

Having confirmed that locus 2 in the RER was subject to coordinate misregulation at smaller window sizes I then went on to confirm the compaction results seen at this locus in other breast cancer cell lines. I used the heat map analysis of expression data for locus 2 to select another ER+ve (LY2) and another ER-ve (MDAMB468) cell line (Figure 4.4.: Yellow box, black arrows). These two cell lines also showed a coordinate and consistent expression phenotype of relative activation in ER+ve cells and repression in ER-ve cells to one another across the locus.

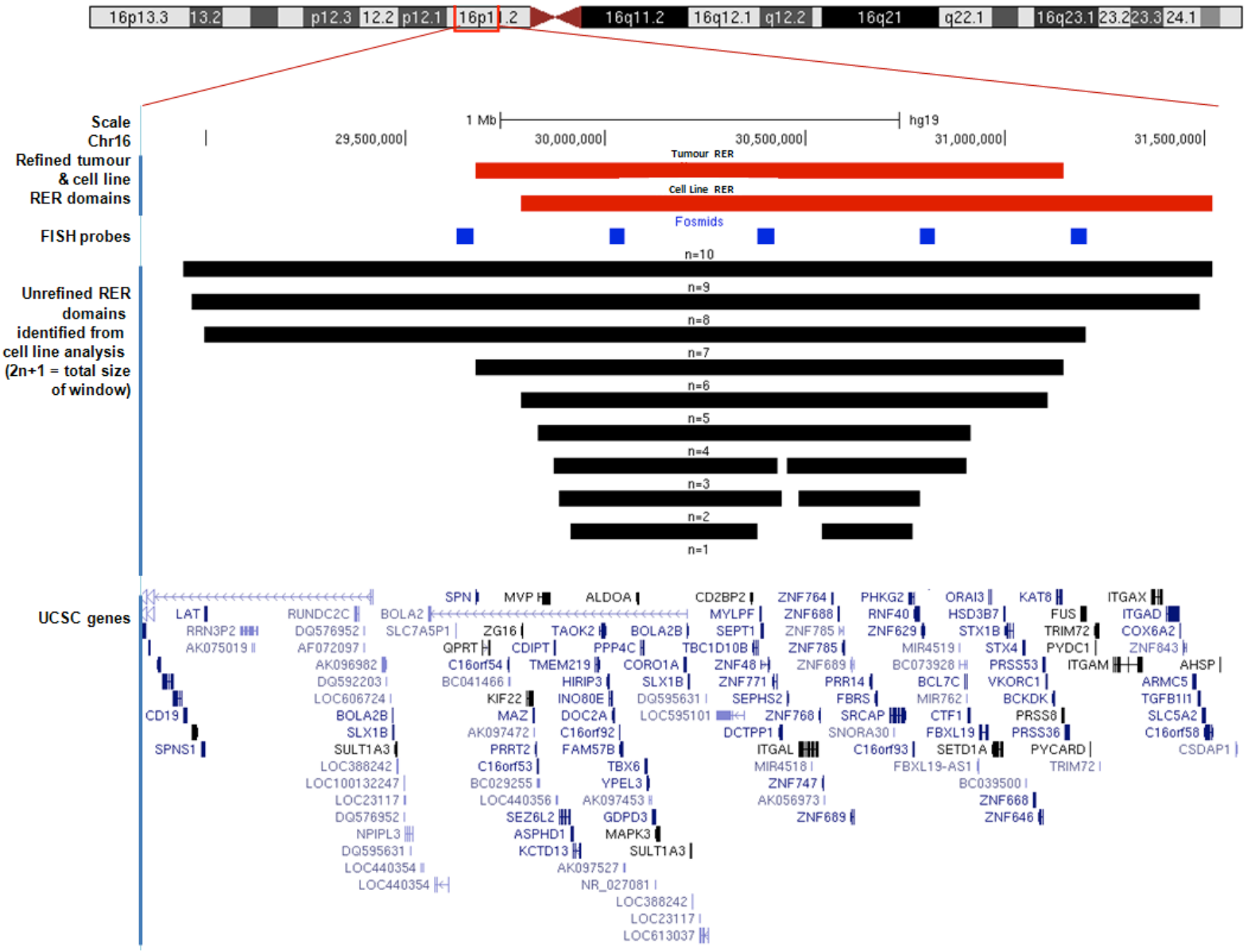

16p13.3
13.2
p12.3
12.2
p12.1
16p1
1.2
16q11.2
16q12.1
q12.2
16q21
q22.1
16q23.1
23.2
23.3
24.1
Scale
Chr16
1 Mb
hg19
29,500,000
30,000,000
30,500,000
31,000,000
31,500,000
Refined tumour & cell line RER domains
Tumour RER
Cell Line RER
FISH probes
Fosmids
Unrefined RER domains identified from cell line analysis (2n+1 = total size of window)
n=10
n=9
n=8
n=7
n=6
n=5
n=4
n=3
n=2
n=1
UCSC genes
SPN
MVP
ALDOA
CD2BP2
ZNF764
PHKG2
ORAI3
KAT8
ITGAX
LAT
RUNDC2C
BOLA2
MYLPF
ZNF688
RNF40
HSD3B7
FUS
ITGAD
RRN3P2
DQ576952
SLC7A5P1
ZG16
TAOK2
BOLA2B
SEPT1
ZNF785
ZNF629
STX1B
TRIM72
COX6A2
AK075019
AF072097
QPRT
CDIPT
PPP4C
TBC1D10B
ZNF785
MIR4519
STX4
PYDC1
ZNF843
AK096982
C16orf54
TMEM219
CORO1A
ZNF48
ZNF689
BC073928
PRSS53
ITGAM
AHSP
DQ592203
BC041466
HIRIP3
SLX1B
ZNF771
PRR14
BCL7C
VKORC1
ARMC5
LOC606724
KIF22
INO80E
DQ595631
SEPHS2
FBRS
MIR762
BCKDK
TGFB1I1
CD19
BOLA2B
MAZ
DOC2A
LOC595101
ZNF768
SRCAP
CTF1
PRSS8
SLC5A2
SLX1B
AK097472
C16orf92
DCTPP1
SNORA30
FBXL19
PRSS36
C16orf58
SPNS1
SULT1A3
PRRT2
FAM57B
ITGAL
C16orf93
SETD1A
PYCARD
CSDAP1
LOC388242
C16orf53
TBX6
MIR4518
FBXL19-AS1
TRIM72
LOC100132247
BC029255
YPEL3
ZNF747
BC039500
LOC23117
LOC440356
AK097453
AK056973
ZNF668
DQ576952
SEZ6L2
GDPD3
ZNF689
ZNF646
NPIPL3
ASPHD1
MAPK3
DQ595631
KCTD13
SULT1A3
LOC440354
AK097527
LOC440354
NR_027081
LOC388242
LOC23117
LOC613037

Figure 4.6: RER analysis at 16p11.2 with decreasing the sliding window size. Ideogram at the top represents chromosome 16 with a red box encompassing the region (p11.2) highlighted below containing the RER domain. The browser tracks represents the location of the RER identified from the analysis of 356 breast tumours (top) or 48 breast cancer cell lines (below). The position of fosmids (blue) used as probes for analysis of the region by FISH are indicated. Below this are browser tracks (black) representing regions derived from the cell line data when varying the number of neighbours in the sliding analysis window, from n=1-10 (where 2n+1 = total window size), before refinement of the region.

The normalised inter-probe FISH distances recapitulated my previous findings of a significantly more decompact chromatin architecture in ER+ve (MCF7, LY2) versus ER-ve breast cancer cells (MDAMB231, MDAMB468) (Figure 4.7). The significance of these differences were assessed using the Wilcox test and are summarised in Table 4.1. Interestingly there was also a significant difference in chromatin compaction between LY2 and MCF7 cells with LY2 appearing to be more decompact at this locus (Figure 4.7; Table 4.1). This may be due to differences in the nuclear size between the cell lines.

Table 4.1: Summary Table of all Wilcox test p-values for cell line analysis of normalised interprobe distances for locus 2 in the RER on chromosome 16p11.2

Cell lines tested		Wilcox test p-value
HMLE	MCF7	0.001
HMLE	LY2	0.000
HMLE	MDAMB361	0.029
HMLE	MDAMB231	0.196
HMLE	MDAMB468	0.071
MCF7	LY2	0.022
MCF7	MDAMB361	0.127
MCF7	MDAMB231	0.026
MCF7	MDAMB468	0.043
LY2	MDAMB361	0.000
LY2	MDAMB231	0.000
LY2	MDAMB468	0.000
MDAMB361	MDAMB231	0.358
MDAMB361	MDAMB468	0.716
MDAMB231	MDAMB468	0.465

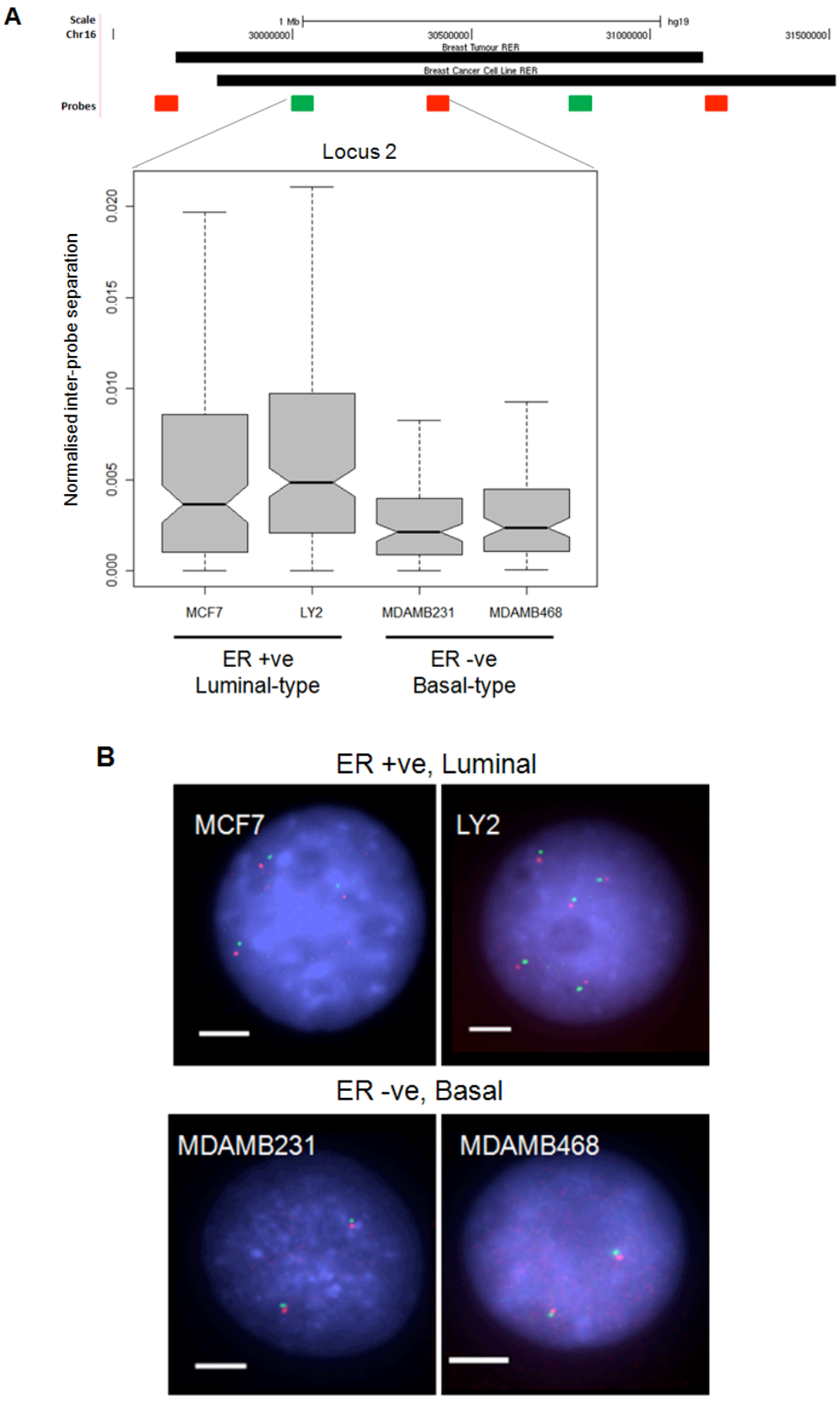


Figure 4.7: Chromatin compaction at locus 2 correlates with breast cancer subtype. (A) As viewed in the UCSC browser with scale and location on chromosome 16p11.2. The top browser track represents the location of the RER identified from the analysis of 356 breast tumours. The bottom browser track represents the location of the RER identified from the analysis of 48 breast

cancer cell lines. The FISH probes browser track (red and green) represents the location of fosmids used as probes for analysis of the region by FISH, dividing the region into 4 loci. n = 45-60 interphase nuclei. and normalised to nuclear area. The distribution of these normalised interprobe distances in two luminal-type ER+ve cell lines (MCF7, LY2) and two basal-type ER-ve cell lines (MDAMB231,MDAMB468) are represented as boxplots below the tracks described. (B) Representative FISH images of nuclei hybridised with probes W12-1754H9 (green) and W12-906G10 (red). DNA is stained with DAPI (blue). Scale bars = 5μm.

The ER+ve cell lines generally had a significant larger nuclear size compared with the ER-ve cell lines (Figure 4.8; Table 4.2). MCF7 cells from these preparations had much larger nuclear area than all other cell lines tested though these cells did not have the largest normalised inter-probe distances at this locus. Interestingly it was the LY2 cells that had the largest normalised inter-probe distances at locus 2 even though they were not the cells with the largest nuclear area they were significantly larger than MDAMB231 they were not larger than MDAMB468 cells. There was also a significant difference between the nuclear area between MDAMB231 and MDAMB468 cells though there was no significant change in the level of chromatin compaction at this locus. This data would indicate therefore that the difference in nuclear size between cell lines does not explain the changes in chromatin compaction seen between the ER+ve and ER-ve cell lines.

Additionally, the two ER+ve cell lines were mostly aneuploidy with three chromosome 16's in MCF7 cells and five in LY2 cells (Figure 4.7B). The precise relationship between ploidy and nuclear area is unknown and therefore the variable ploidy seen in these cell lines could account for the changes seen in chromatin compaction between the two subtypes. To eliminate this possibility I selected a negative control region to study by FISH, at which I would not expect to see chromatin compaction differences between breast cancer sub-types.

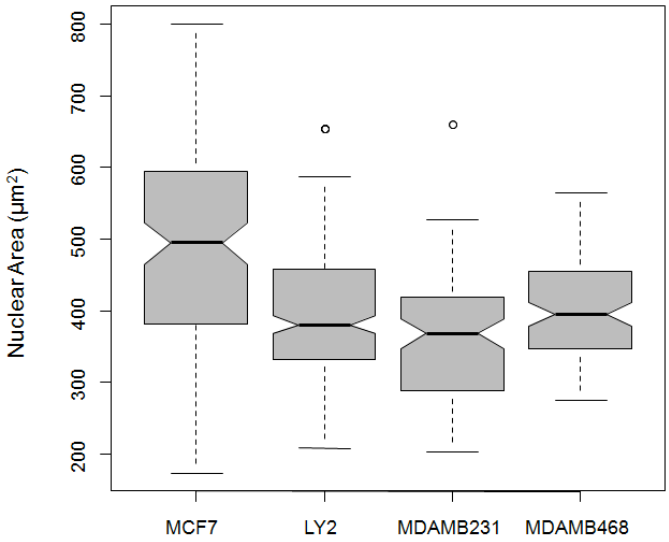


Figure 4.8: Nuclear area in breast cancer cell lines. Box plots of the nuclear area (μm^2) distribution across all nuclei tested in the four breast cancer cell lines as determined from DAPI staining. Data from FISH analysis of locus 2 was used for this analysis (n=40-60).

Table 4.2: Summary Table of all Wilcox test p-values for cell line analysis of nuclear area using FISH data for locus 2 analysis on chromosome 16p11.2

Cell lines tested		Wilcox p-value
MCF7	LY2	0.000
MCF7	MDAMB231	0.000
MCF7	MDAMB468	0.000
LY2	MDAMB231	0.043
LY2	MDAMB468	0.110
MDAMB231	MDAMB468	0.002

4.2.4: Selection of a negative control region

To ensure that there are no systematic variations in the chromatin compaction between the cell lines used in the above analysis of the RER on chromosome 16p11.2 it was necessary to investigate a control locus. Previously in the Bickmore lab the *β-globin* locus on chromosome 11 has been used as a FISH control region as these genes are expressed in a tissue specific manner in erythroid cells. The *β-globin* locus is also surrounded by olfactory receptor genes which were generally thought to be expressed exclusively in the olfactory sensory neurons, although recent expression profiling has shown that they are expressed in a wide variety of human tissues (Flegel et al., 2013). I used a pair of probes 533 kb apart that span the *β-globin* locus to examine the two luminal-type ER+ve cell lines (MCF7, LY2) and the two basal-type ER-ve cell lines (MDAMB231, MDAMB468). (Figure 4.9).

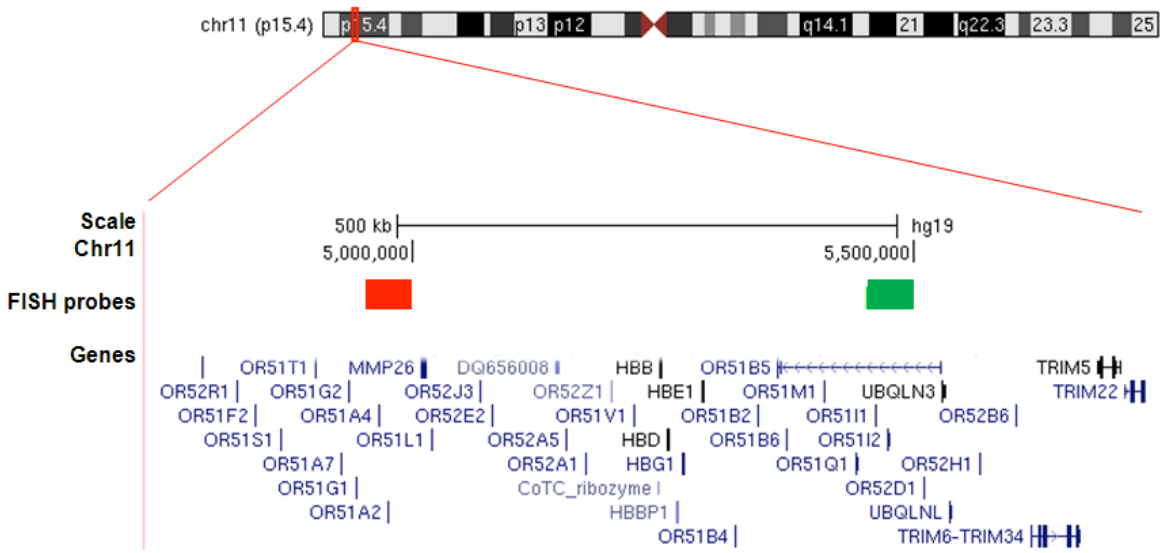


Figure 4.9: A FISH control locus on chromosome 11p15.4 encompassing the β-globin and olfactory receptor loci. As viewed in the UCSC browser. Ideogram at the top represents chromosome 11 with a red box encompassing the region (p15.4) highlighted below containing the control domain. The FISH probes browser track (red: W12-528M6; green:W12-2033J5) represents the location of fosmids used as probes for analysis of the region by FISH. Below the browser track are all genes within the coordinates for the red box as determined by UCSC.

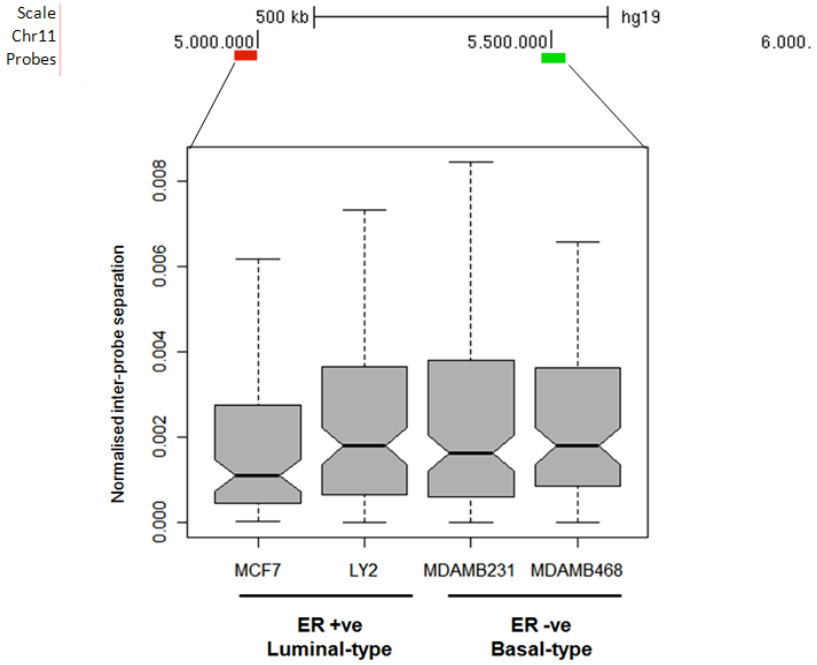
There was no significant difference between the distribution of normalised inter-probe distances for this region between LY2, MDAMB231 and MDAMB468 cell lines (Figure 4.10; Table 4.3). However the region appears to be more compact in MCF7 cells than in all other cell lines analysed (Figure 4.10; Table 4.3). This is not what I would have expected if there was a systematic change in the chromatin compaction between the cell lines tested for locus 2 on chromosome 16p11.2 – I would have expected a mirror of the results shown in Figure 4.7 and Table 4.1.

Table 4.3: Summary Table of all Wilcox test p-values for cell line analysis of normalised interprobe distances the control region on chromosome 11

Cell lines tested		Wilcox test p-value
MCF7	LY2	0.023
MCF7	MDAMB231	0.056
MCF7	MDAMB468	0.010
LY2	MDAMB231	0.665
LY2	MDAMB468	0.790
MDAMB231	MDAMB468	0.493

LY2 and MDAMB231 cells both have three copies of chromosome 11 whereas the MCF7 and MDAMB468 cells have two (Figure 4.10B). This difference in ploidy appears to have no bearing on the chromatin compaction status of the control locus as although both MCF7 and MDAMB468 have normal ploidy for this chromosomes there is a significant difference in the chromatin compaction with a more decompact structure for the MDAM468 cells akin to that seen for the cell lines LY2, MDAMB231 (Figure 4.10A, Table 4.3).

A



B

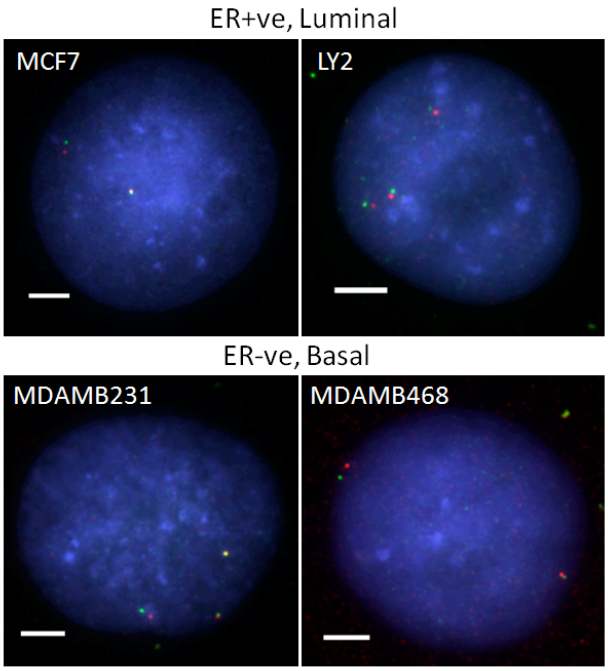


Figure 4.10: Chromatin compaction at control region on chromosome 11p15.4. (A) As viewed in the UCSC browser with scale and location on chromosome 11. The probes browser track (red and green) represents the location of fosmids used as probes for analysis of the region by FISH. For this control locus the interprobe distance was assayed in 45-60 interphase nuclei and normalised to nuclear area. The distribution of these normalised interprobe distances in two luminal-type ER+ve cell lines (MCF7, LY2) and two basal-type ER-ve cell lines (MDAMB231,MDAMB468) are represented as boxplots. (B) Representative images of FISH nuclei for all cell lines hybridised with probes W12-2033J5 (green) and W12-528M6 (red). DNA is stained with DAPI (blue). Scale bars = 5μm.

I then set out to identify a control region that was as similar to the region of subtype specific compaction change (locus 2) as possible, and located on the same chromosome (16). The control region identified was 1.4 Mb upstream of the start of locus 2 and was not in any of the RER domains I previously identified (Figure 4.11).

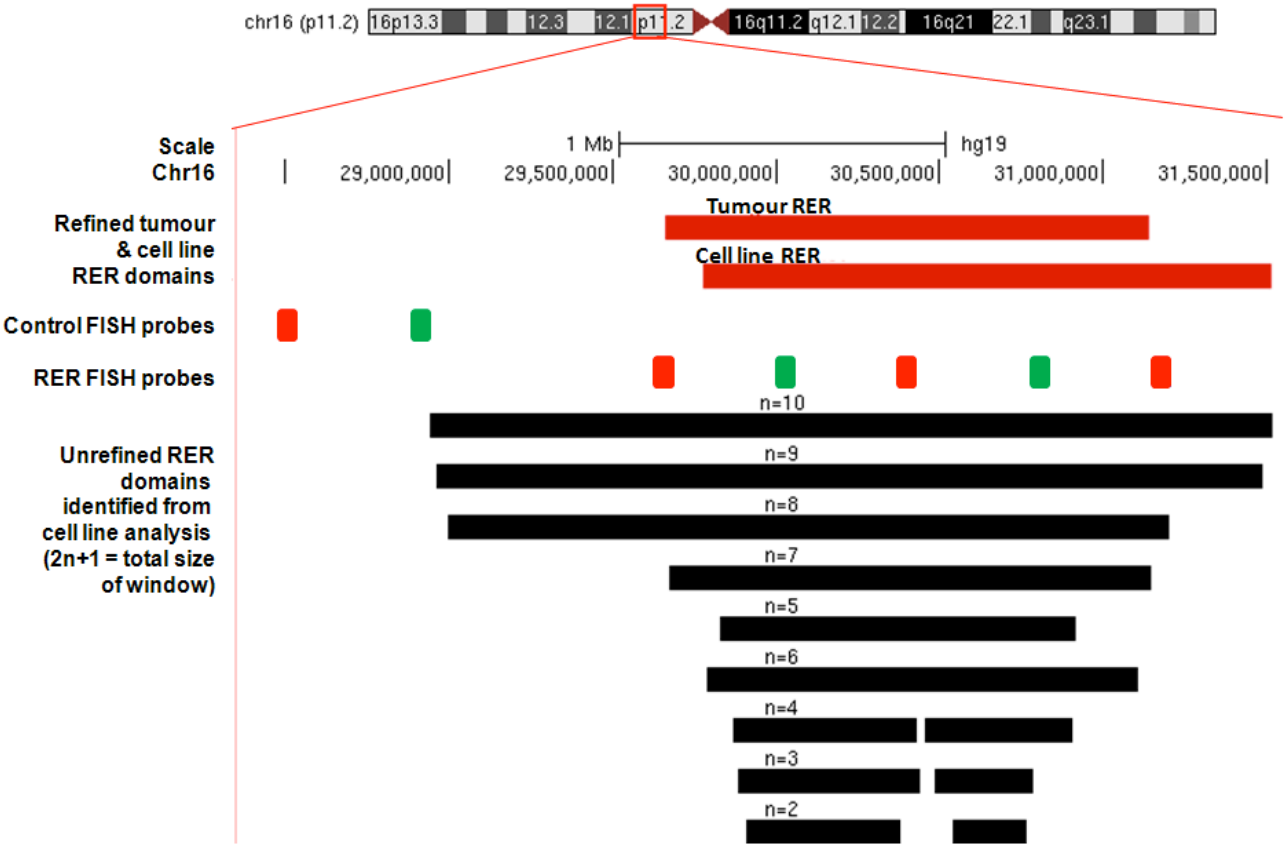


Figure 4.11: UCSC genome browser tracks showing the location of control FISH probes in relation to the RER on chromosome 16p11.2. Ideogram at the top represents chromosome 16 with a red box encompassing the region (p11.2) highlighted below containing the RER domain. The top browser track represents the location of the RER identified from the analysis of 356 breast tumours. The bottom browser track represents the location of the RER identified from the analysis of 48 breast cancer cell lines. The control FISH probes browser track (red and green) shows the start and end point of the control region selected. The RER FISH probes track below (red and green) represents the location of fosmids used as probes for analysis of the RER region by FISH. Below this are browser tracks (black) representing regions derived from the cell line data when varying the number of neighbours in the sliding analysis window, from n=1-10 (where 2n+1 = total window size), before refinement of the region.

Figure 4.12 shows all genes within the 380 bp region selected as a FISH control domain on chromosome 16p11.2.

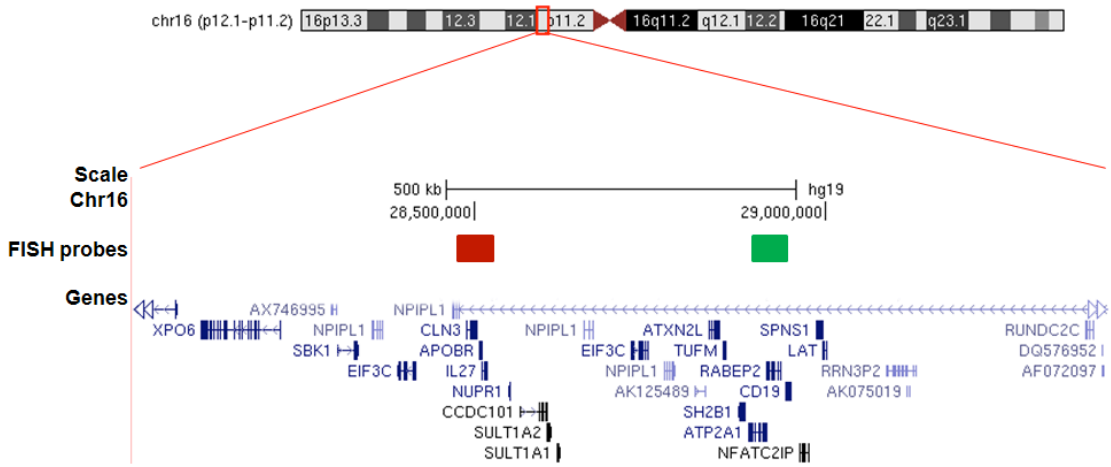


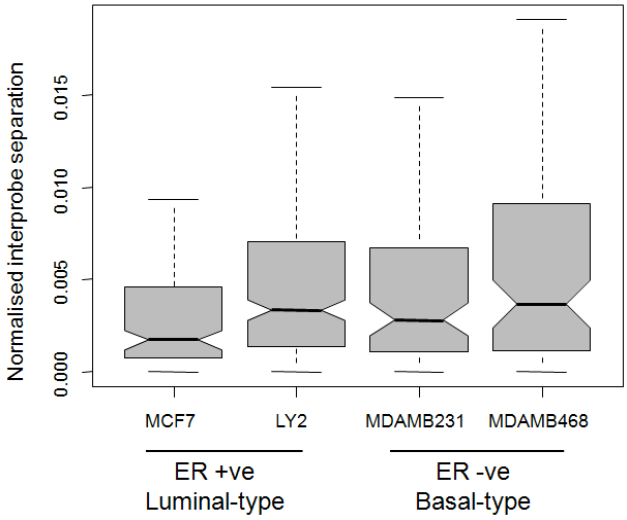
Figure 4.12: Control locus on chromosome 16p11.2. The FISH probes browser track (red: WI2-3081O2; green:WI2-2889N9) represents the location of fosmids used as probes for analysis of the region by FISH. Below the browser track are all genes within the coordinates for the red box as determined by UCSC.

FISH using this control locus in the two ER+ve (MCF7, LY2) and ER-ve (MDAMB231, MDAMB468) breast cancer cell lines show that the differences in nuclear area and ploidy between these cell lines was not responsible for the changes seen in chromatin compaction seen in the 16p locus 2 RER (Figure 4.13, Table 4.4). As for the region encompassing the *β-globin* locus, the chromatin at this second control region was most compact in the MCF7 cells when compared to all other cell lines. No other significant differences in chromatin compaction were seen between the other cell lines.

Table 4.4: Summary Table of all Wilcox test p-values for cell line analysis of normalised interprobe distances for the control locus located adjacent to the RER on chromosome 16p11.2

Cell lines tested		Wilcox test p-value
MCF7	LY2	0.000
MCF7	MDAMB231	0.026
MCF7	MDAMB468	0.001
LY2	MDAMB231	0.252
LY2	MDAMB468	0.914
MDAMB231	MDAMB468	0.305

A



B

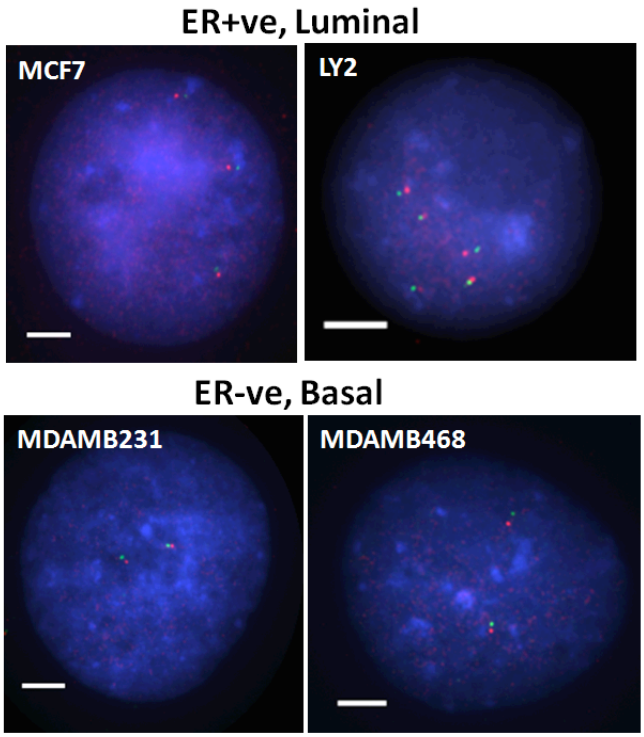


Figure 4.13: Chromatin compaction at control region on chromosome 16p11.2. The interprobe distance was assayed in 50-60 interphase nuclei and normalised to nuclear area. The distribution of these normalised interprobe distances in two luminal-type ER+ve cell lines (MCF7, LY2) and two basal-type ER-ve cell lines (MDAMB231,MDAMB468) are represented as boxplots. (B) Representative images of FISH nuclei for all cell lines hybridised with probes WI2-2889N9 (green) and WI2-3081O2 (red). DNA is stained with DAPI (blue). Scale bars = 5μm.

4.2.5: Chromatin compaction in a normal breast cell line

I had established that the changes in chromatin compaction at locus 2 of the RER on chromosome 16p11.2 in breast cancer cell lines are related to tumour subtype. I detected a relative chromatin decompaction in ER+ve cell lines at this locus compared with ER-ve cell lines and this is correlated with gene expression signature. To show that altered chromatin compaction was indeed related to coregulated gene expression I selected, from the heat map analysis of expression z scores for locus 2, a cell line where there was not a coordinate expression pattern across the entire locus (an "intermediate expression phenotype").

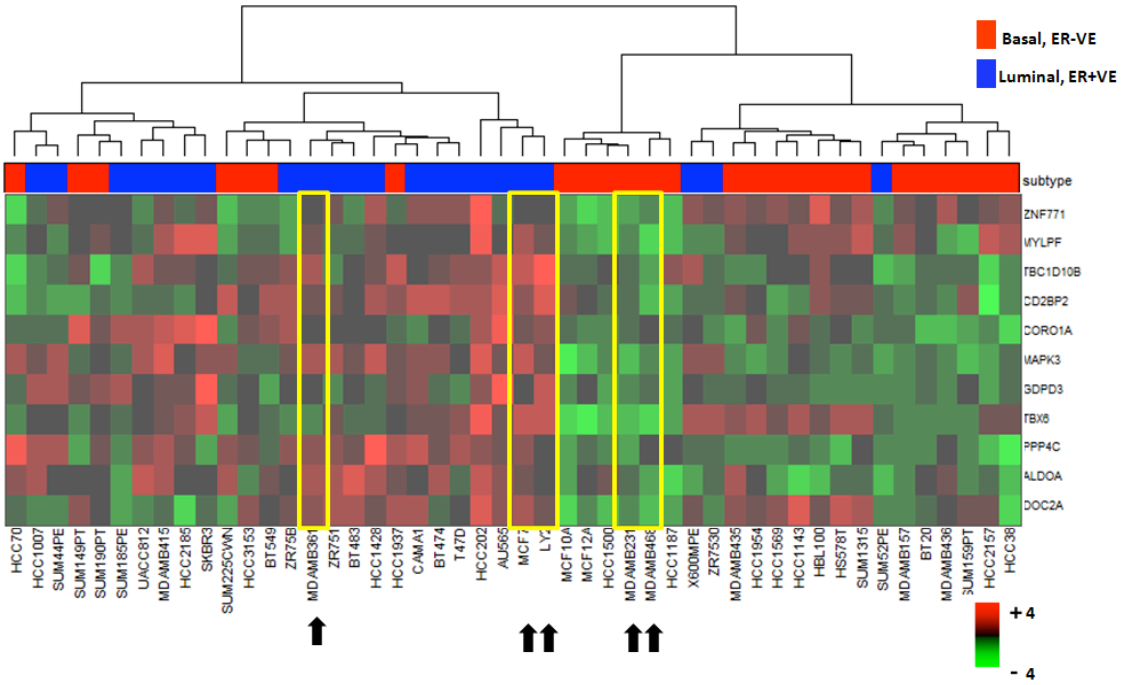
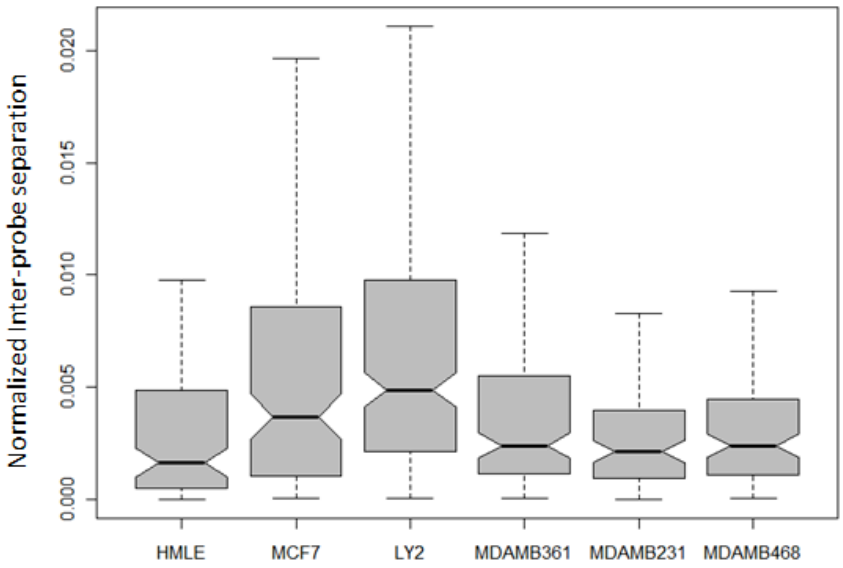


Figure 4.14: Selecting a cell line with an intermediate gene signature in 16p11.2 RER. Unsupervised cluster analysis of locus 2 for 48 breast cell lines from Neve et al., 2006. The heat map is vertically clustered using Euclidean clustering (shown as dendogram) to indicate the similarity of sample scores with each other. Based on subtype information from Neve et al., 2006 each sample is identified by a color-coded matrix below the dendrogram: basal (red), luminal (blue) The legend shows z-scores where red/green indicate an increase/decrease in gene expression relative to the universal mean for each gene. The values were mean centred and the colours scaled from the max (+4) & min (-4) standard deviations. This enables relative comparison of scores between samples. Yellow box and arrows indicate cell lines used for FISH.

Consistent with its gene expression profile, the normalised inter-probe distances established by FISH in the MDAMB361 cell line indicated a chromatin compction level intermediate to that in ER+ve (MCF7, LY2) and ER-ve (MDAMB231,MDAMB468) cells (Figure 4.15). The distribution of inter-probe distances for MDAMB361 cells was not significantly different to that of any of the other breast cancer cell lines except LY2 which was the most decompact cell line overall as well as when compared to the other ER+ve cell line MCF7 (Table 4.1).

Given that gene expression in the 16p11.2 RER is thought to be aberrantly upregulated in tumours compared to normal breast tissue (Figure 4.1A) I next wanted to examine the chromatin compaction status at locus 2 of a normal breast cell line. To do this, I carried out FISH on fixed preparations of a non-transformed immortalized human mammary epithelial cell line HMLE obtained from Elad Katz (Breakthrough Breast Cancer) and developed in the Weinberg lab (Elenbaas et al., 2001; Rangarajan et al., 2004); (Ma et al., 2010). Analysis of normalised inter-probe distances showed that locus 2 is more compact in chromatin structure in HMLE cells than in all ER+ve cell lines tested (MCF7, LY2, MDAMB361), but not compared to ER-ve cell lines (MDAMB231, MDAMB468) (Figure 4.15; Table 4.1). This was not an altogether surprising finding given that the largest incidence of tumours tend to be ER+ve and that the analysis of tumour expression data against normal tissues showed an upregulation in tumour tissues (Figure 4.1A).

A



B

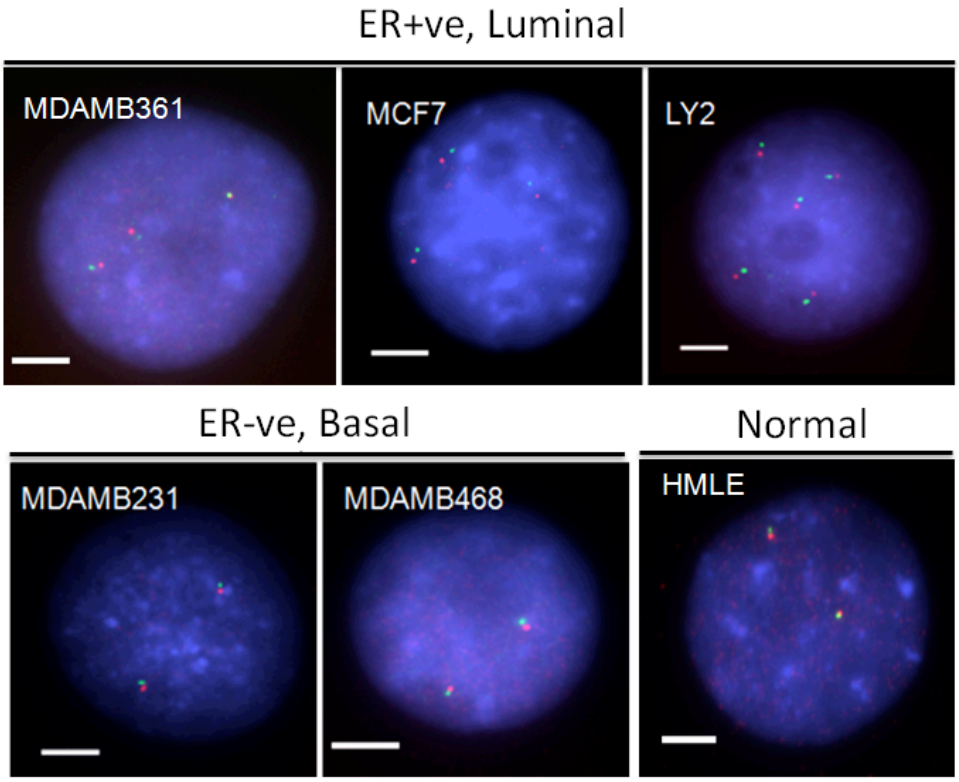


Figure 4.15: Chromatin compaction at locus 2 correlates with breast cancer subtype. (A) As viewed in the UCSC browser with scale and location on chromosome 16p11.2. The top browser track represents the location of the RER identified from the analysis of 356 breast tumours. The bottom browser track represents the location of the RER identified from the analysis of 48 breast cancer cell lines. The FISH probes browser track (red and green) represents the location of fosmids used as probes for analysis of the region by FISH, dividing the region into 4 loci. For Locus 2 the interprobe distance was assayed in 45-60 interphase nuclei and normalised

to nuclear area. The distribution of these normalised interprobe distances in three luminal-type ER+ve cell lines (MCF7, LY2, MDAMB361), two basal-type ER-ve cell lines (MDAMB231,MDAMB468) and a normal breast cell line (HMLE) are represented as boxplots below the tracks described. (B) Representative images of FISH nuclei for all cell lines hybridised with probes W12- 1754H9(green) and W12-906G10 (red). DNA is stained with DAPI (blue). Scale bars = 5μm.

4.2.6: Confirmation of chromatin compaction status in tumour tissue sections

Having confirmed, for a candidate RER, changes in chromatin compaction that correlated with changes in gene expression *in vitro* I then wanted to examine this *ex vivo*. To do so I obtained breast tumour tissue sections and normal breast tissue by collaboration with Jeremy Thomas (Consultant Pathologist at Edinburgh Breast Unit). Usage of tumour material was approved by the Lothian Research Ethics Committee (08/S1101/41) and obtained under the auspices of Experimental Cancer Medicine Centre programme (Edinburgh).

3D FISH analysis using probes for locus 2 (RER on chromosome 16p11.2) on tissue sections from an ER+ve tumour, an ER-ve tumour and normal breast tissue confirmed previous findings from cell-line based work. The chromatin at this locus was at its most compact in the normal tissue with the most significant decompaction observed in ER+ve tissue (Figure 4.16; Table 4.5). Moreover there was no significant difference in chromatin compaction between normal tissue and ER-ve tumour at this locus (Table 4.5). There was however, as expected from the cell line data, a significant difference in the chromatin compaction between ER+ve and ER-ve tumours for this locus contained in the RER on chromosome 16p11.2 (Figure 4.16; Table 4.5).

Table 4.5: Summary Table of all Wilcox test p-values for breast tumour and normal tissue analysis of normalised interprobe distances for the RER (locus 2) located on chromosome 16p11.2

Tissues tested		Wilcox test p-value
Normal	ER+ve	0.000
Normal	ER-ve	0.240
ER+ve	ER-ve	0.004

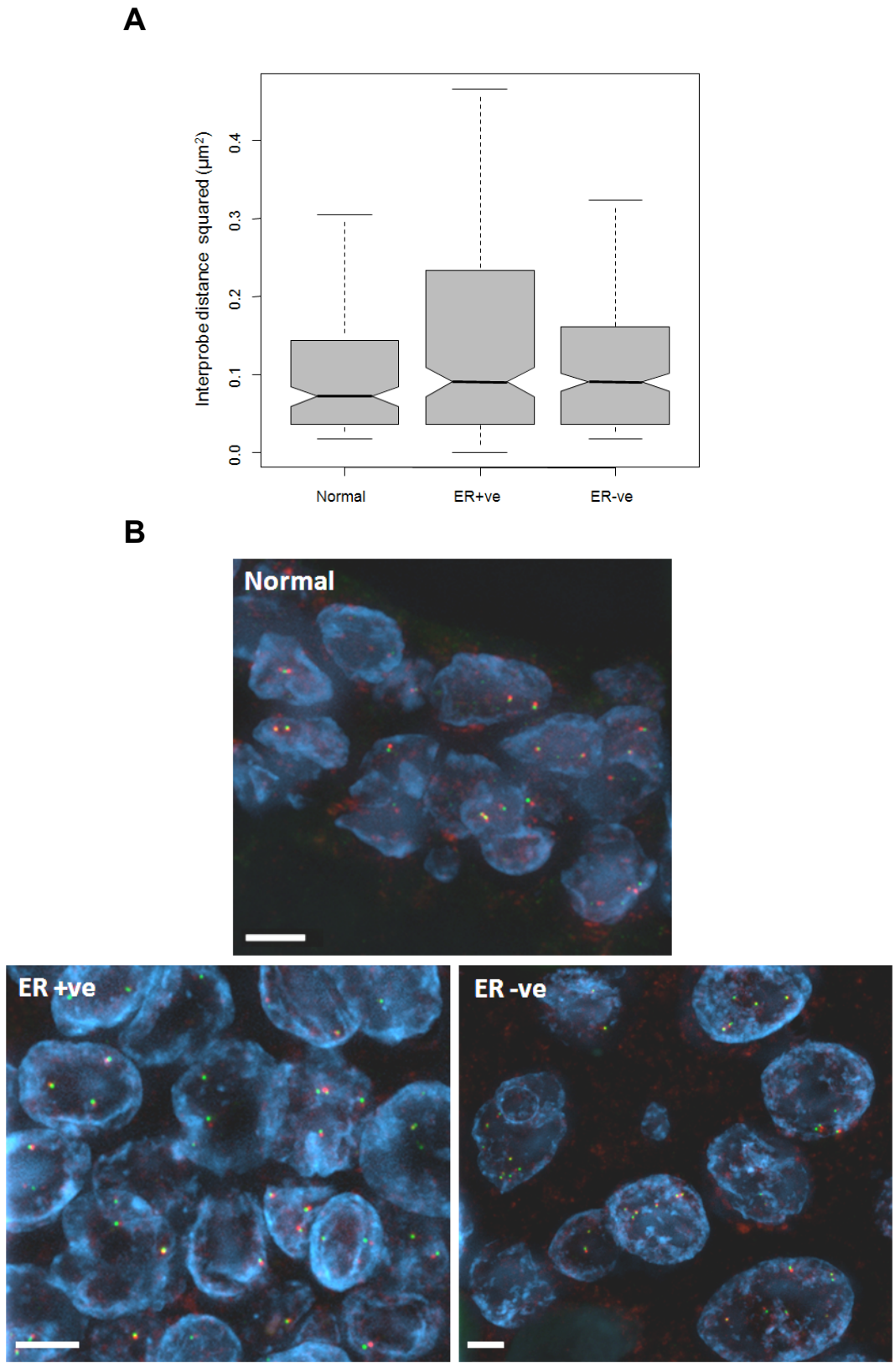


Figure 4.16: Chromatin compaction at locus 2 of RER correlates with breast cancer subtype in tumour tissues by 3D FISH. (A). The squared interprobe distances for 250-300 pairs of FISH probes was assayed in normal breast tissue, an ER+ve type tumour tissue and an ER-ve type tumour tissue. The distribution of these interprobe distances are represented as boxplots.

(B) Representative snapshot images of tissue sections used for FISH analysis with probes W12-1754H9 (green) and W12-906G10 (red). DNA is stained with DAPI (blue). Scale bars = 5μm.

4.2.7: Data for chromatin compaction follows a random-walk model

A random walk model of the behaviour of higher order chromatin has been identified using FISH data with pairs of probes at different genomic intervals along the linear DNA molecule (van den Engh et al., 1992). With probe separations of 100kb to 1.5 Mb there is a linear relationship between the mean inter-probe distances (squared) and the genomic distance that fits a "random walk" model with the chromatin behaving like a flexible polymer (van den Engh et al., 1992). This model was also suggested by further work using probes for 150kb to 190Mb intervals on chromosomes 4, 5 and 19 and modelling of the large-scale chromatin geometry (Sachs et al., 1995; Yokota et al., 1995). This analysis produced a "random walk/giant loop" model of large-scale chromatin organisation whereby at megabase intervals the chromatin is organised in loop structures which are fixed on a flexible backbone that shows random walk behaviour (Sachs et al., 1995).

In this model interprobe distance conform to a Rayleigh distribution and are characterised by a standard deviation/mean ratio of ~0.52 and a median/mean ratio of ~0.94. Similar data distributions have been seen in the Bickmore lab for many regions of the mouse and human genomes in a wide variety of cell types and in tissue sections (Chambeyron and Bickmore 2004, Gilbert et al., 2004, Williamson et al., 2012). In this chapter the distribution of observed distances for each of the cell lines I tested also fitted this distribution, with the exception of the MCF7 (locus 1) and HMLE (locus 2) cell lines (Table 4.6). Therefore, with these exceptions, at the genomic distances tested in this chapter the chromatin fibre between two probes follows a random walk model. The deviation from this random walk model by the MCF7 and HMLE cells suggests an alternative behaviour of the chromatin at locus 1 and 2 in these cell lines.

Table 4.6: Summary of all cell-line based 2D FISH analyses of inter-probe distances

Cell Line	RER domain	FISH locus	Probe 1	Probe 2	Mean inter-probe distance (µm)	Standard Deviation /Mean	Median /Mean
MCF7	Chr 16p11.2	Locus 1	WI2-1584N4	W12-1754H9	0.569	0.824	0.746
MCF7	Chr 16p11.2	Locus 2	W12-1754H9	W12-906G10	0.793	0.603	0.941
MCF7	Chr 16p11.2	Locus 3	W12-906G10	W12-497E18	0.913	0.621	0.884
MCF7	Chr 16p11.2	Locus 4	W12-497E18	W12-2222M4	0.884	0.761	0.801
MDAMB231	Chr 16p11.2	Locus 1	WI2-1584N4	W12-1754H9	0.517	0.720	0.834
MDAMB231	Chr 16p11.2	Locus 2	W12-1754H9	W12-906G10	0.582	0.664	0.833
MDAMB231	Chr 16p11.2	Locus 3	W12-906G10	W12-497E18	0.786	0.665	0.948
MDAMB231	Chr 16p11.2	Locus 4	W12-497E18	W12-2222M4	0.620	0.659	0.864
LY2	Chr 16p11.2	Locus 2	W12-1754H9	W12-906G10	0.852	0.618	0.914
MDAMB468	Chr 16p11.2	Locus 2	W12-1754H9	W12-906G10	0.607	0.535	0.882
HMLE	Chr 16p11.2	Locus 2	W12-1754H9	W12-906G10	0.492	0.826	0.767
MDAMB361	Chr 16p11.2	Locus 2	W12-1754H9	W12-906G10	0.670	0.684	0.824
MCF7	Chr 11p15.4	Control	W12-2033J5	W12-528M6	0.561	0.601	0.853
LY2	Chr 11p15.4	Control	W12-2033J5	W12-528M6	0.556	0.562	0.950
MDAMB231	Chr 11p15.4	Control	W12-2033J5	W12-528M6	0.478	0.608	0.925
MDAMB468	Chr 11p15.4	Control	W12-2033J5	W12-528M6	0.599	0.574	0.909
MCF7	Chr 16p11.2	Control	WI2-2889N9	WI2-3081O2	0.714	0.650	0.833
LY2	Chr 16p11.2	Control	WI2-2889N9	WI2-3081O2	0.641	0.607	0.883
MDAMB231	Chr 16p11.2	Control	WI2-2889N9	WI2-3081O2	0.658	0.707	0.825
MDAMB468	Chr 16p11.2	Control	WI2-2889N9	WI2-3081O2	0.668	0.641	0.889
MCF7	Chr 16p11.2	Locus 2	W12-1754H9	W12-906G10	0.749	0.682	0.842
MDAMB231	Chr 16p11.2	Locus 2	W12-1754H9	W12-906G10	0.568	0.666	0.830
MCF7	Chr 16p11.2	Control	WI2-2889N9	WI2-3081O2	0.720	0.625	0.828
MDAMB231	Chr 16p11.2	Control	WI2-2889N9	WI2-3081O2	0.570	0.575	0.940

In normal breast and tumour tissue samples the observed distances by FISH also conformed to a Rayleigh distribution for each of the cell lines tested (Table 4.7) thereby indicating that the *ex vivo* 3D FISH analyses also followed the random walk model of chromatin behaviour. This was also true of the normal breast tissue tested by 3D FISH unlike the normal breast epithelial cell line (HMLE).

Table 4.7 Summary of tissue based 3D FISH analyses of inter-probe distances

Tissue	RER domain	FISH locus	Probe 1	Probe 2	Mean inter-probe distance	Standard Deviation	Median /Mean
ER+ve	Chr 16p11.2	Locus 2	W12-1754H9	W12-906G10	0.337	0.508	0.890
ER-ve	Chr 16p11.2	Locus 2	W12-1754H9	W12-906G10	0.381	0.553	0.787
Normal	Chr 16p11.2	Locus 2	W12-1754H9	W12-906G10	0.300	0.508	0.894

4.2.8: RER localisation with regard to the nuclear periphery

As discussed in chapter 1 (sections 1.2.4.4. and 1.3.4) the spatial organisation of chromosomes in the nucleus influences gene expression, with the nuclear periphery associated with transcriptional repression. Differences in the radial position of genes in the nucleus have been reported that can reliably detect cancerous tissues (Meaburn et al., 2009). I therefore set out to determine whether or not there were differences in the radial positioning of the RER on chromosome 16p11.2. To quantify radial positioning, I compared the relative proportion of hybridization signals from the RER on chromosome 16p11.2 across five shells of equal area eroded from the centre (shell 1) through to the periphery (shell 5) in 40-60 nuclei from each cell line (Figure 4.17).

Chromosome 16 is normally located in the centremost erosion shells of nucleus (S Boyle et al., 2001) therefore it was not unexpected to find that most of the hybridisation signals were located in shells 1-2 (Figure 4.17). Given that the ER+ve cell lines (MCF7, LY2) show a higher expression signature and more decompact chromatin structure at this RER compared to ER-ve cell lines (MDAMB231, MDAMB468), I examined whether the ER+ve cell lines had more hybrisation signals in the centremost erosion shells of the nucleus. Indeed whilst this appears to be the case for the centremost erosion shell 1, overall the difference in the distribution of hybridisation signals across erosion shells was only significant between ER+ve MCF7 cells and the ER-ve MDAMB231 cells (Table 4.8, Figure 4.17).

Table 4.8 Summary of statistics testing changes in the radial positioning of 16p11.2

Cell lines tested		Fisher exact test (p-value)
MCF7	MDAMB231	0.015
MCF7	MDAMB468	0.289
MCF7	LY2	0.238
LY2	MDAMB231	0.261
LY2	MDAMB468	0.708
MDAMB231	MDAMB468	0.338

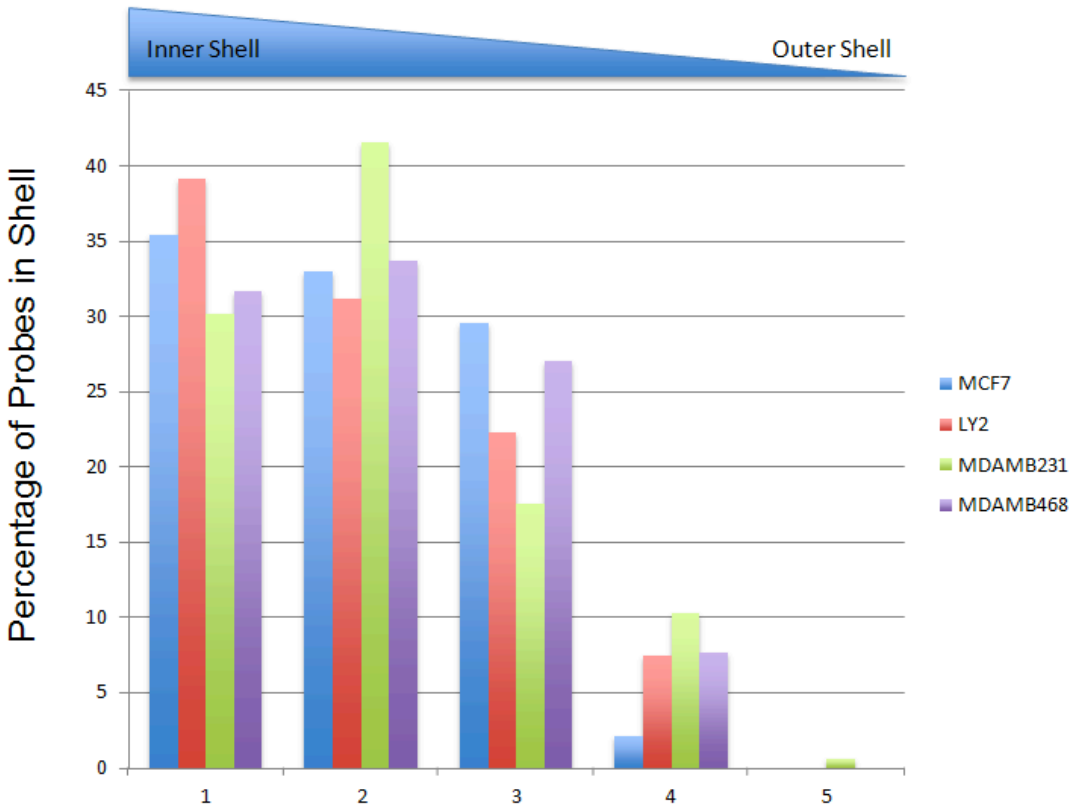


Figure 4.17: Eroding the nucleus into shells of equal nuclear area to determine the localisation of RER on chromosome 16p11.2 Shown are the percentage of FISH probe signals for this region that occurred in each of five shells of equal area with increasing shell number indicating shells away from the centre of the nucleus.

4.3: Discussion

In the previous chapter I identified at least 26 regions in the breast cancer genome that show aberrant coordinate gene expression potentially due to epigenetic regulation. These were consistent between breast cancer cell line and tumour datasets and were denoted Regions of Epigenetic Regulation (RERs). The most interesting part of this analysis was that these RERs were not all subject to the LRES phenomena described in previous studies, rather they showed a variety of expression phenotypes that included aberrantly "activated" and "repressed" expression when comparing breast tumours to normal tissue.

In this chapter I selected a RER on chromosome 16p11.2 to investigate the mechanism of dysregulation that appeared in analyses from both tumour and cell line data and that was also differentially expressed between different subtypes of breast cancer. Given that previous reports have shown multiple epigenetic changes in LRES domains I hypothesized that an alteration in the higher order chromatin organisation may have a role to play in misregulating such large regions of the genome in cancer. The selected RER domain also contained genes (*SPN*, *QPRT, ZG16, KIF22, MAZ, PRRT2, MVP, SEZ6L2, ASPHD1, KCTD13, TMEM219, TAOK2, INO80E, DOC2A*) in a region previously shown to be involved in estrogen-mediated repression by DNA looping in breast cancer (Hsu et al., 2010) which also made it an interesting candidate for investigating chromatin structure. This region is aberrantly upregulated in breast cancer.

Having refined the domain of RER by dividing it into 4 loci for FISH and using smaller sliding windows for computational analyses, I investigated chromatin decompaction as a mechanism of the aberrant activation of gene expression at this RER on chromosome 16p11.2 (denoted locus 2). Previous work in the Bickmore lab has shown by FISH that gene-rich transcriptionally active loci are in a decondensed chromatin structure in the nucleus which is thought to be a permissive environment for transcription to take place by allowing access to the genome (Gilbert et al., 2004). My findings in this chapter also showed that where the chromatin for locus 2 was decompacted in ER+ve breast cancer cell lines and tumour tissue, transcription was activated compared to normal and ER-ve

cell lines and tissue. Control FISH loci were used to confirm that the changes observed were not systemic.

Whereas, these findings are consistent with what is expected from previous studies that link active gene expression and chromatin decompaction, it is not what would have been expected based on data from (Hsu et al., 2010). Hsu et al describe a "large repressive zone" in breast cancer cells that is encompassed by my original RER on chromosome 16p11.2 (specifically FISH locus 1) and that involves 14 gene promoters in a DNA loop structure thought to act as a physical repressive barrier to transcription (Figure 4.19A, green boxes). It was suggested that in normal cells this is a transient estrogen mediated structure but is a permanent feature of cancer cells (Hsu et al., 2010). I would therefore have expected the ER+ve subtype of breast cancer cells to be significantly more compact at loci 1. However, this was not the case and the region as a whole was also not repressed in tumours relative to normal tissue (Figure 4.19B: p-value = 1.403e-09). Heat map analysis of the cell line expression data shows that there is not an obvious subtype-specific coordinated expression pattern at locus 1 taken by itself (Figure 4.19C). Looking at the cell lines used for FISH this is particularly true of the MCF7 cell line. Therefore it is not surprising that upon refinement of the region using smaller window sizes this locus was lost as part of the 16p11.2 RER.

I also observed a re-localisation of the 16p11.2 region more toward the nuclear centre in ER+ve MCF7 cells compared with ER-ve MDAMB231 cells. This was not however a significant effect when comparing ER+ve LY2 and ER-ve MDAMB468 cell lines. It's possible that this is due to the highly rearranged nature of chromosomes in the cancer cell lines, especially for the LY2 cell line, which would likely effect positioning.

Overall, the data in this chapter suggest that this refined RER domain of 16p11.2 is aberrantly upregulated in breast cancer due to an overall decompaction of chromatin of the locus and possibly its repositioning towards the centre of the nucleus - perhaps to hubs containing transcriptional machinery required for active gene expression (Sutherland and Bickmore, 2009). In chapter 5 I further investigate the mechanism underlying this chromatin organisation by focussing on the role of the ER.

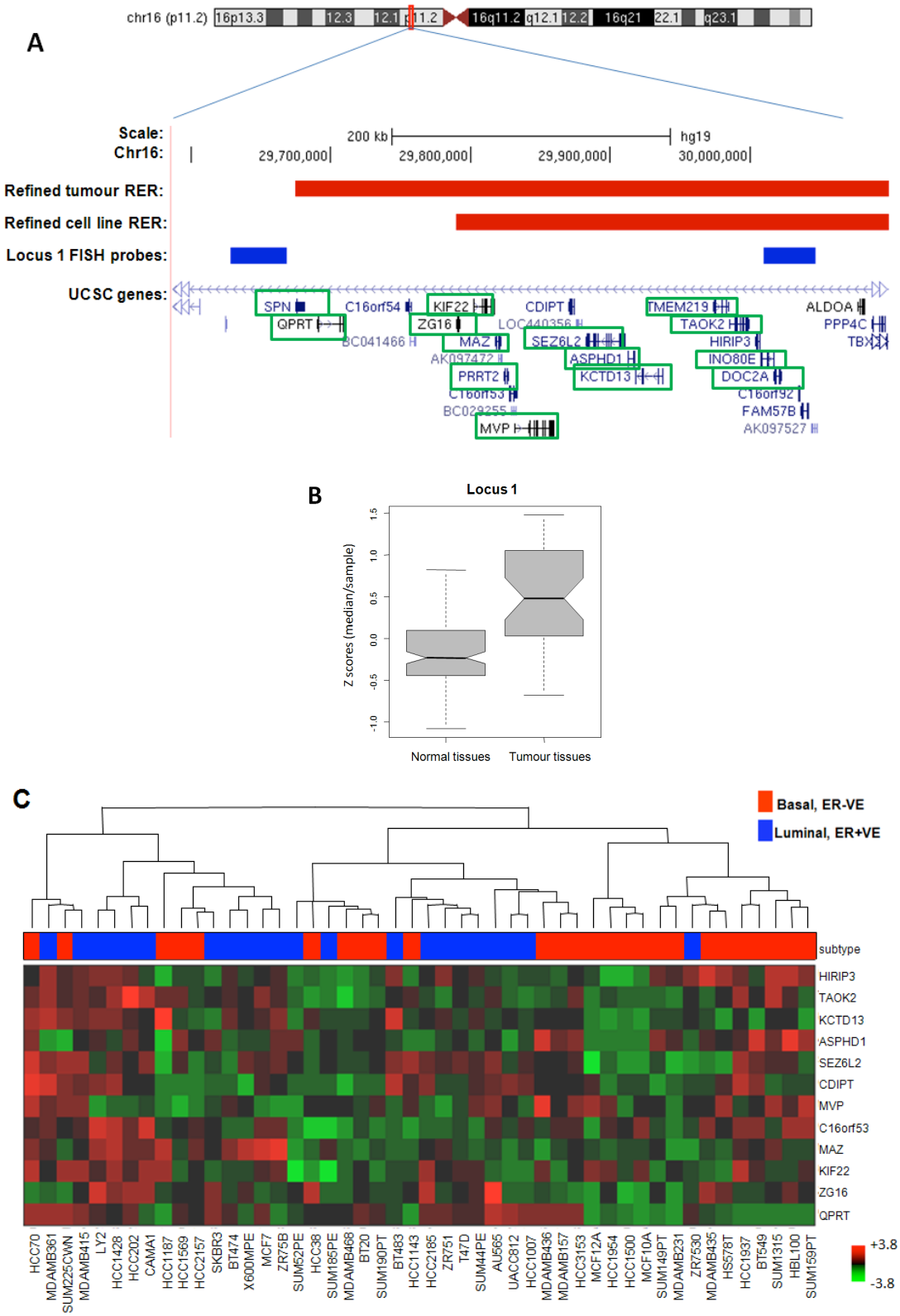


Figure 4.19: Assessing locus 1 of the RER on chromosome 16p11.2. (A) As viewed in the UCSC browser. Ideogram at the top represents chromosome 16 with a red box encompassing the region (p11.2) highlighted below containing RER domain in tumours and cell lines (red) and locus 21 FISH probes below (blue). At bottom are all genes within the coordinates for the red box

as determined by UCSC. Boxed in green are the 14 genes involved in DNA looping by Hsu et al 2010. (B) Analysis of gene expression changes in tumours relative to normal tissue. Mean centred z scores were calculated for all genes in the region. The median z score for each sample was used to to plot the distribution of these scores as boxplots and enabling relative comparison of scores between samples. (C) Unsupervised cluster analysis of locus 1 for 48 breast cell lines from Neve et al., 2006. The heat map is vertically clustered using Euclidean clustering (shown as dendogram) to indicate the similarity of sample scores with each other. Based on subtype information from Neve et al., 2006 each sample is identified by a color-coded matrix below the dendrogram: basal (red), luminal (blue) The legend shows z-scores where red/green indicate an increase/decrease in gene expression relative to the universal mean for each gene. The values were mean centred and the colours scaled from the max (+3.81) & min (-3.81) standard deviations. This enables relative comparison of scores between samples.

Chapter 5

Changes in chromatin compaction at locus 2 of the 16p11.2 RER domain are estrogen-mediated

5.1: Introduction

In the previous chapter I began investigating a proposed RER domain at 16p11.2. This RER was of particular interest as it was aberrantly upregulated in breast tumours and showed a differential expression phenotype across ER+ve and ER-ve tumour subtypes. The RER was further refined by division into four sub-regions which were assayed by FISH to assess chromatin compaction. At sub-region 2 chromatin decompaction was observed that correlated with breast cancer subtype and gene expression signature. Further computational work based on efforts described in chapter 3 showed that the RER domain could be further delineated into a much smaller RER domain that correlated with FISH locus 2.

Previous studies looking for large regions of coordinated gene misregulation in cancer have observed changes in DNA methylation and histone modifications present at particular genes, though these are not uniformally distributed through the regions. Based upon this premise I set out to investigate alterations in chromatin architecture as a plausible explanation for the misregulation of large regions of contiguous genes. In this part of the thesis I investigate the hypothesis that the alteration in chromatin packing at the RER on chromosome 16p11.2 is mediated by the estrogen receptor.

5.2: Results

5.2.1: Sub-region 2 of the 16p11.2 RER contains genes bound by ERα

The changes in chromatin architecture described in Chapter 4 showed subtype specificity in relation to the ER status of the cell lines and tumours. As described in the Introduction to this thesis (Section 1.2.4.6.) the estrogen receptor has been reported to have a role in chromatin organisation. I therefore wanted to see what ER binding sites there were in locus 2 of the 16p11.2 where I had observed the subtype specific chromatin organisation changes. Using the ChIA-PET visualisation browser (http://cms1.gis.a-star.edu.sggure) I was able to identify a number of ER binding sites in the locus that overlapped with H3K4me3, RNA polymerase II and FoxA1 binding in estrogen induced MCF7 cells (Figure 5.1) (Fullwood et al., 2009). Moreover from the ChIA-PET ERα data the region was not enriched in complex or duplex intra-chromosomal interactions in response to estrogen treatment (Figure 5.1 – browser track 5).

By extrapolating from their complex intra-chromosomal associations assayed by ChIA-PET, Fullwood et al (2009) postulate that there are estrogen-mediated chromatin loop structures, that allow interacting 'anchor' genes to be actively transcribed in close proximity to one another. This is conceptually similar to the transcription factory model. Contrary to this, Hsu et al (2010) showed that estrogen-mediated cross-linked associations bring together multiple gene promoters linked to their transcriptional repression. However as discussed at the end of Chapter 4 this analysis was not supported by my own findings for this reported domain (FISH locus 1).

From the ChIA-PET data it was evident that locus 2 of the 16p11.2 RER domain was largely devoid of the complex estrogen-induced intra-chromosomal interactions in ER+ve MCF7 cells (Figure 5.1) that are described in Fullwood et al (2009). This is in keeping with my observation that ER+ve cell lines and tumours are more decompact at this locus than those that are ER-ve, since presumably complex interactions and looping within the locus should result in a chromatin conformation that appears more compact in ER+ve

cells/tumours by FISH (Fullwood et al 2009). To confirm this, and to discover the impact of the ER with regard to compaction at locus 2, I decided to investigate the behaviour of this locus before and after estrogen-treatment of ER+ve and ER-ve cells.

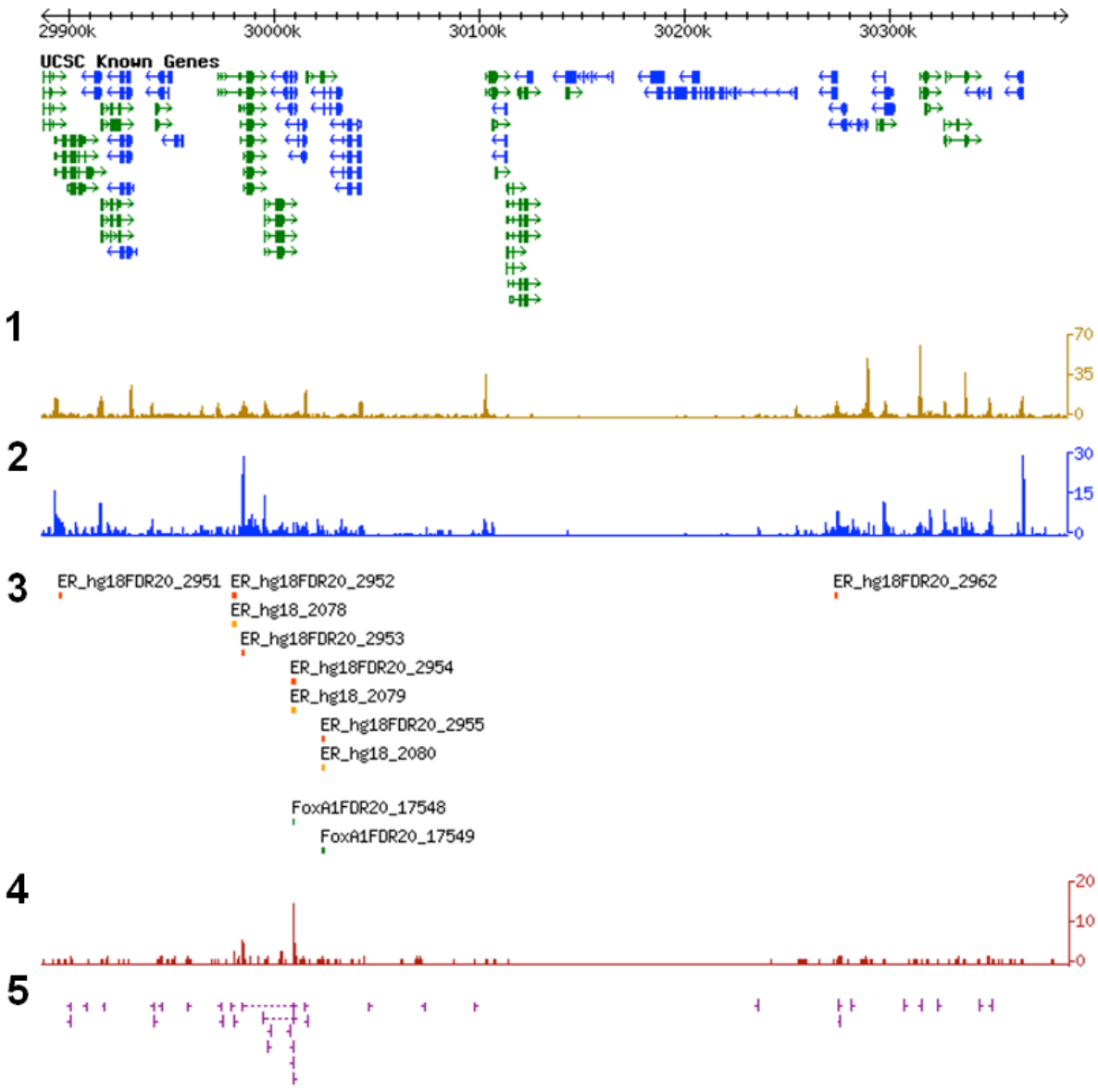


Figure 5.1: Locus 2 of RER 16p11.2 in estrogen-induced MCF7 cells. Screenshot taken from the ChIA-PET visualation browser: http://cms1.gis.a-star.edu.sg (Fullwood et al 2009) using hg18 coordinates for FISH locus 2. ChIA-PET browser tracks: 1, H3K4me3 ChIP-Seq; 2, RNAPII ChIP-Seq; 3, ERα (orange) and FoxA1 ChIP-chip (green) (Lupien et al., 2008); 4, ERα ChIA-PET density; 5, Intra-chromosome Interactions.

5.2.2: The affect of estrogen on chromatin compaction

Having concluded that there were ER binding sites in locus 2 of RER 16p11.2 (Figure 5.1) I wanted to test my hypothesis that the changes in higher order chromatin architecture were mediated by estrogen and the ER. To do this I cultured the ER+ve breast cancer cell line MCF7, and ER-ve cell line MDAMB231, in phenol-free media with fetal calf serum that had been stripped of all endogenous hormones (see Methods). After three consecutive days in this media under "starvation" conditions I treated the cells with 100nM of 17β-estradiol (hereby referred to as estrogen) for 24 hours. These conditions have been shown in time-course experiments to activate high level expression of estrogen-regulated genes in MCF7 cells (Karp et al., 2007; Tynan et al., 2004).

Cells were then fixed to make preparations for FISH, from normal media (N), starved media (S) and starved media with the reintroduced estrogen (E) and chromatin compaction was assayed by measuring the physical distance between probes (W12-1754H9 and W12-906G10) flanking locus 2 of RER 16p11.2 (Figure 4.2). Inter-probe distances from 50-60 images were converted from pixels to microns and were then were squared (d^2). The mean d^2 was used to estimate the level of chromatin compaction under the different conditions. The observed inter-probe distances were also normalised to the radius squared (d^2/r^2) to account for differences in the nuclear area between samples. As before, significant changes in chromatin compaction were assessed using the Wilcox test with a cut off of $p<0.05$ (summarised in Table 5.1).

Analysis of the data for ER+ve MCF7 cells showed that upon hormone deprivation the chromatin at locus 2 compacts significantly and that this effect can be rescued by addition of estrogen to the starved media (Figure 5.2 and Table 5.1). This is consistent with the idea that the decompact chromatin status at locus 2 for ER+ve cells is mediated by estrogen itself.

Table 5.1: Summary of statistics for detecting significance of estrogen-mediated changes

Cell lines tested		Locus tested	Wilcox test p-value
MCF7 (Normal)	MCF7 (Starved)	locus 2 RER	0.002
MCF7 (Normal)	MCF7 (Starved + Estrogen)	locus 2 RER	0.520
MCF7 (Starved)	MCF7 (Starved + Estrogen)	locus 2 RER	0.003
MCF7 (Normal)	MCF7 (Starved)	control	0.000
MCF7 (Normal)	MCF7 (Starved + Estrogen)	control	0.000
MCF7 (Starved)	MCF7 (Starved + Estrogen)	control	0.944
MDAMB231 (Normal)	MDAMB231 (Starved)	locus 2 RER	0.406
MDAMB231 (Normal)	MDAMB231 (Starved + Estrogen)	locus 2 RER	0.766
MDAMB231 (Starved)	MDAMB231 (Starved + Estrogen)	locus 2 RER	0.324
MDAMB231 (Normal)	MDAMB231 (Starved)	control	0.540
MDAMB231 (Normal)	MDAMB231 (Starved + Estrogen)	control	0.477
MDAMB231 (Starved)	MDAMB231 (Starved + Estrogen)	control	0.160

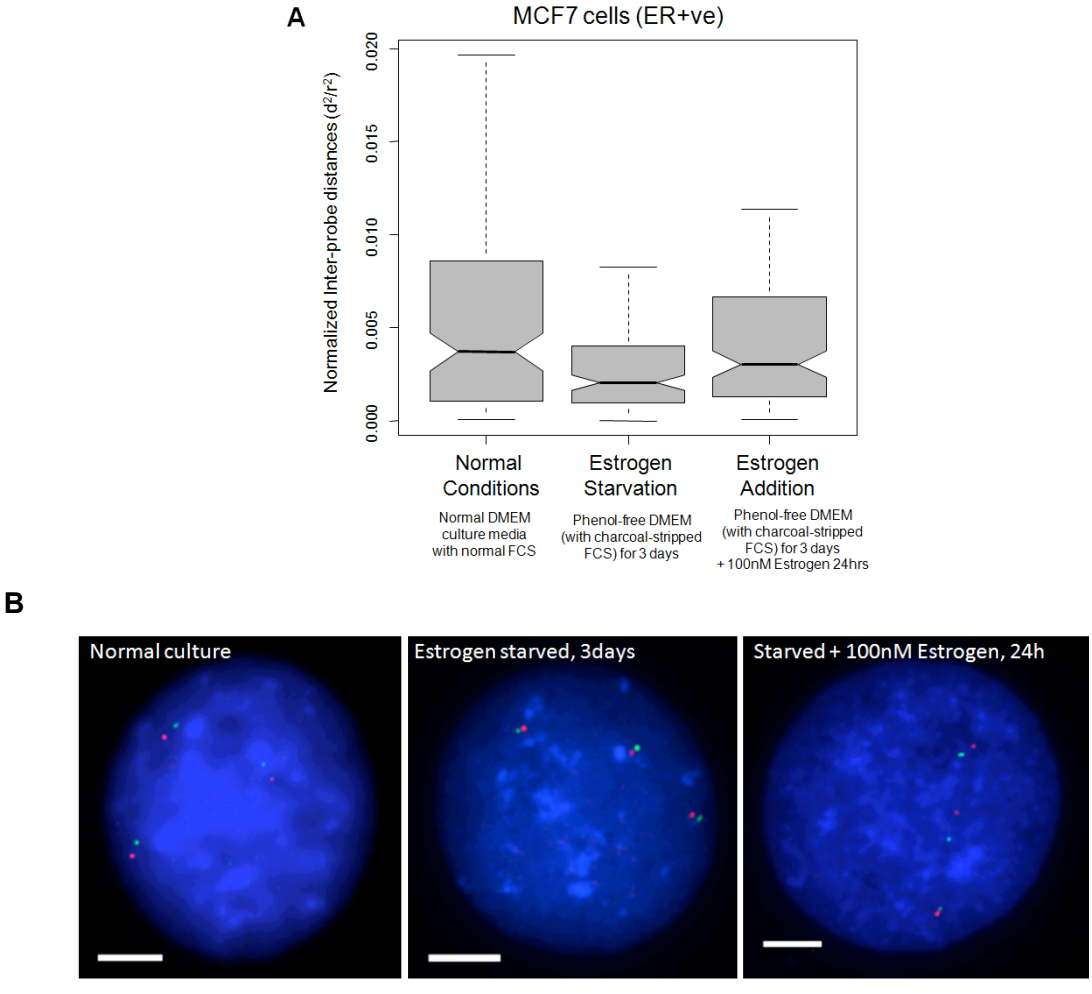
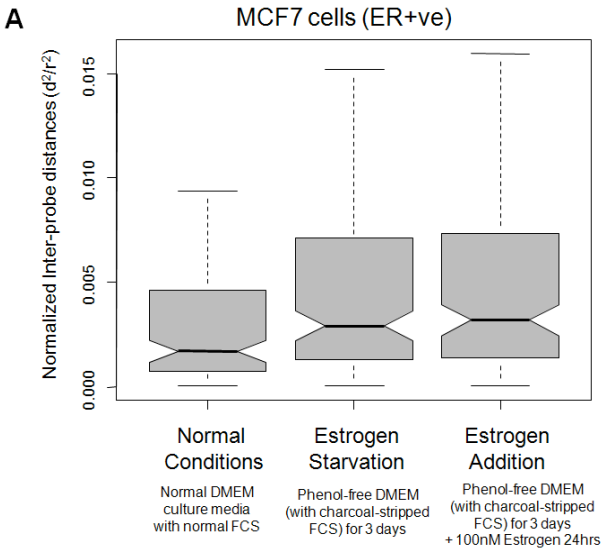


Figure 5.2: Effect of estrogen on chromatin compaction at locus 2 in MCF7 cells. (A) Box plots to show the distribution of FISH interprobe distances (d) normalised to nuclear area (d^2/r^2) in normal media, hormone-deprived media (3 days) and hormone deprived media (3 days) +100nM estrogen (24h). N= 50-60 nuclei (B) Representative images (for each culture condition) of FISH

nuclei hybridised with probes W12-1754H9 (green) and W12-906G10 (red). DNA is stained with DAPI (blue). Scale bars = 5μm.

To ensure there were no systemic changes in gross chromatin compaction across the genome upon these culture conditions I used the FISH control region on chromosome 16p11.2 adjacent to the RER domain as designed in the previous chapter (Figure 5.3 and Table 5.1). Under normal culture conditions the MCF7 cells were at their most compact. Upon hormone deprivation there was a significant decompaction of the control locus that remained decompact after the reintroduction of estrogen to the cells. This latter result suggests that at this control locus estrogen itself does not have an effect on chromatin compaction levels though charcoal-stripping of the serum, which removes hormones and growth factors, does.



B

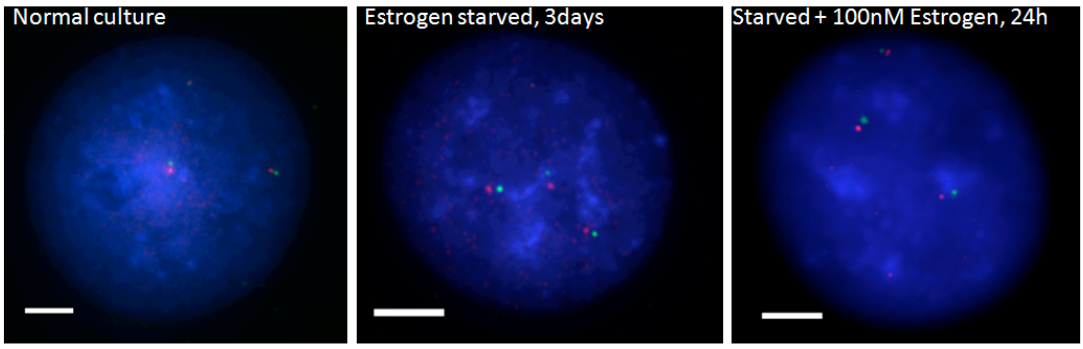


Figure 5.3: Effect of estrogen on chromatin compaction at control locus on 16p11.2 in MCF7 cells. (A) The interprobe distance (d) was assayed in 50-60 interphase nuclei and normalised to nuclear area (d^2/r^2). The distribution of these normalised interprobe distances in normal media, hormone-deprived media (3 days) and hormone derprived media (3 days) +100nM

estrogen (24h) are shown as boxplots. (B) Representative images (for each culture condition) of FISH nuclei hybridised with probes W12-1754H9 (green) and W12-906G10 (red). DNA is stained with DAPI (blue). Scale bars = 5μm.

To show that the changes in chromatin compaction at locus 2 were mediated by estrogen acting through the estrogen receptor, FISH was used to examine chromatin compaction during hormone deprivation and subsequent estrogen addition, in the ER-ve cell line MDAMB231. This showed that chromatin at locus 2 shows no significant change in chromatin compaction after hormone deprivation and after the reintroduction of estrogen in this cell line (Figure 5.4 and Table 5.1).

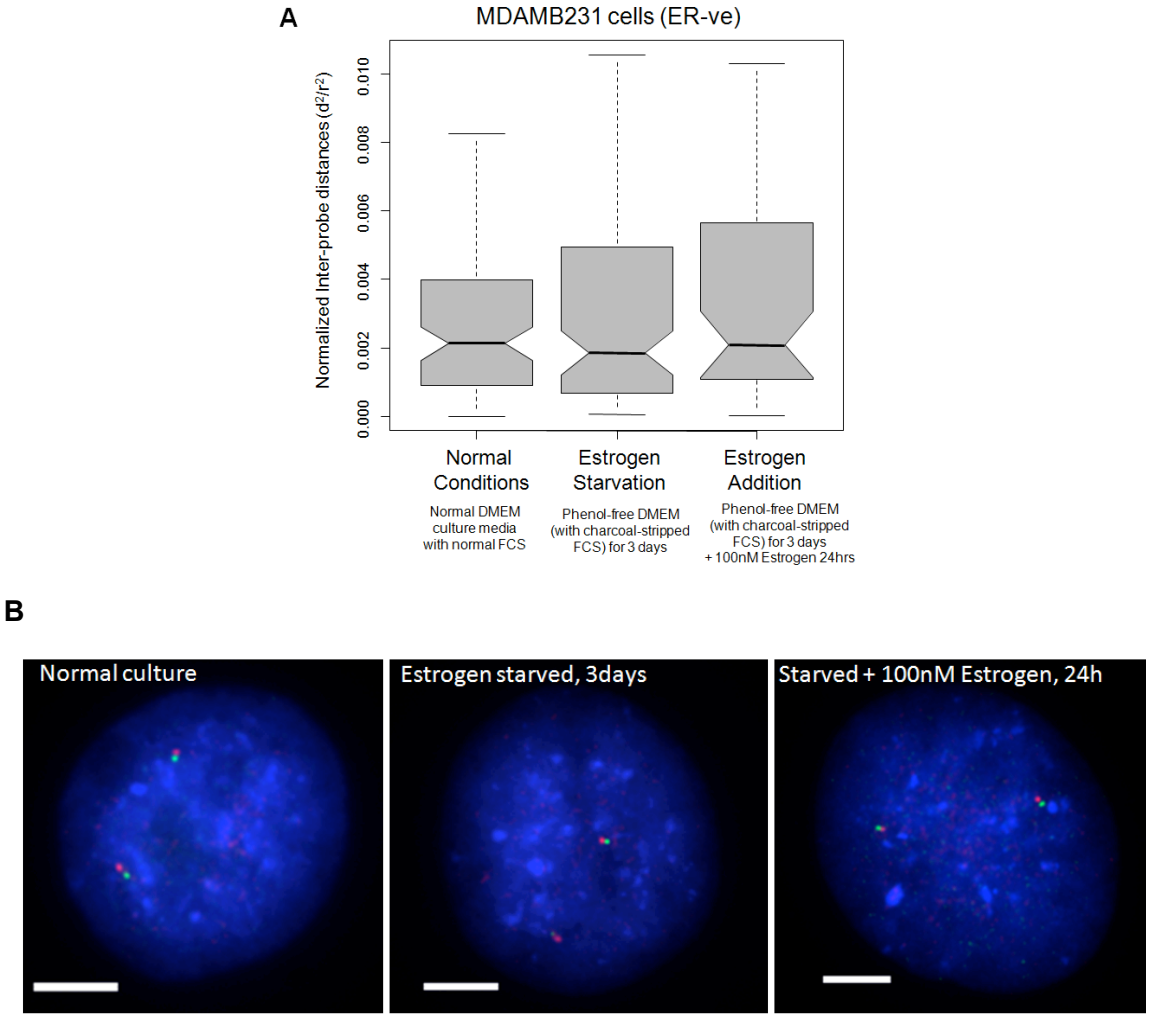
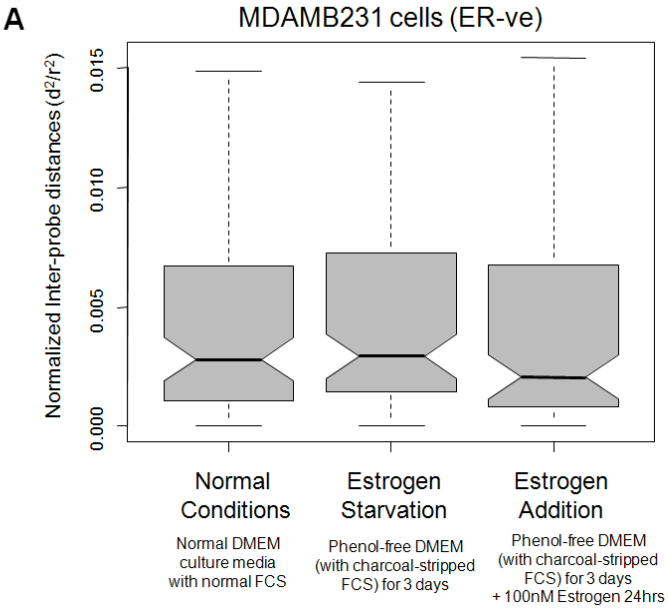


Figure 5.4: Effect of estrogen on chromatin compaction at locus 2 in MDAMB231 cells (A) The interprobe distance (d) was assayed in 50-60 interphase nuclei and normalised to nuclear area (d^2/r^2). The distribution of these normalised interprobe distances in normal media, hormone-deprived media (3 days) and hormone derprived media (3 days) +100nM estrogen (24h) are shown as boxplots. (B) Representative images (for each culture condition) of FISH nuclei hybridised with probes W12-1754H9 (green) and W12-906G10 (red). DNA is stained with DAPI (blue). Scale bars = 5μm.

As for MCF7 cells, the control locus designed adjacent to the RER on chromosome 16p11.2 was assayed for each of the three culture conditions in MDAMB231 cells. This locus showed no change in the level of chromatin packing between each of the conditions (Figure 5.5 and Table 5.1). Therefore I conclude that the changes in chromatin compaction observed in ER+ve MCF7 cells for the RER domain are mediated by estrogen.



B

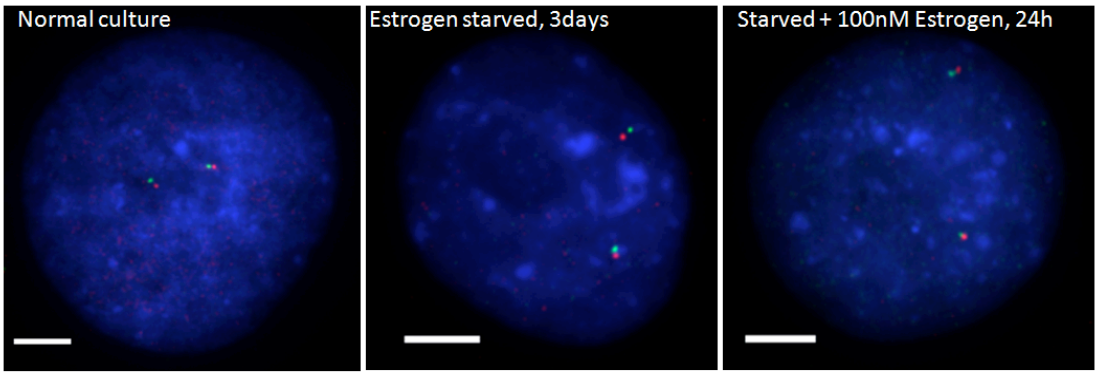


Figure 5.5: Effect of estrogen on chromatin compaction at control locus on 16p11.2 in MDAMB231 cells. (A) The interprobe distance (d) was assayed in 50-60 interphase nuclei and normalised to nuclear area (d^2/r^2). The distribution of these normalised interprobe distances in normal media, hormone-deprived media (3 days) and hormone derprived media (3 days) +100nM estrogen (24h) are shown as boxplots. (B) Representative images (for each culture condition) of FISH nuclei hybridised with probes W12-1754H9 (green) and W12-906G10 (red). DNA is stained with DAPI (blue). Scale bars = 5μm.

As discussed in the previous results chapter the behaviour of higher order chromatin follows a "random walk/giant loop" model where at megabase intervals the chromatin is organised in loop structures which are fixed on a flexible backbone that shows random walk behaviour (Sachs et al., 1995). In this chapter the distribution of observed distances for each FISH analysis conformed to a Rayleigh distribution (i.e. the standard deviation/mean ~0.52 and the median/mean ~0.94) (Table 5.2). Therefore at the genomic distances tested in this chapter the chromatin fibre between two probes follows the random walk model in the nucleus expected as described by previous reports (Sachs et al.,1995;Yokota et al., 1995).

Table 5.2: Summary for Rayleigh Distribution data in estrogen experiments

Conditions	Cell Line	Location	FISH locus	Probe 1	Probe 2	Mean inter-probe distance (µm)	Standard Deviation /Mean	Median /Mean
Normal Media	MCF7	Chr 16p11.2	Locus 2	W12-1754H9	W12-906G10	0.793	0.603	0.941
Normal Media	MDAMB231	Chr 16p11.2	Locus 2	W12-1754H9	W12-906G10	0.582	0.664	0.833
Normal Media	MCF7	Chr 16p11.2	Control	WI2-2889N9	WI2-3081O2	0.714	0.650	0.833
Normal Media	MDAMB231	Chr 16p11.2	Control	WI2-2889N9	WI2-3081O2	0.658	0.707	0.825
Starved Media	MCF7	Chr 16p11.2	Locus 2	W12-1754H9	W12-906G10	0.749	0.682	0.842
Starved Media	MDAMB231	Chr 16p11.2	Locus 2	W12-1754H9	W12-906G10	0.568	0.666	0.830
Starved Media	MCF7	Chr 16p11.2	Control	WI2-2889N9	WI2-3081O2	0.720	0.625	0.828
Starved Media	MDAMB231	Chr 16p11.2	Control	WI2-2889N9	WI2-3081O2	0.570	0.575	0.940
Starved + 100nM Estrogen	MCF7	Chr 16p11.2	Locus 2	W12-1754H9	W12-906G10	0.728	0.628	0.901
Starved + 100nM Estrogen	MDAMB231	Chr 16p11.2	Locus 2	W12-1754H9	W12-906G10	0.450	0.652	0.812
Starved + 100nM Estrogen	MCF7	Chr 16p11.2	Control	WI2-2889N9	WI2-3081O2	0.785	0.744	0.821
Starved + 100nM Estrogen	MDAMB231	Chr 16p11.2	Control	WI2-2889N9	WI2-3081O2	0.641	0.675	0.804

5.2.3: An ER-dependent re-localisation of locus 2 in the nucleus

In the previous chapter I found that the 16p11.2 region was more central in the nucleus in ER+ve MCF7 cells than ER-ve MDAMB231 cells. Given that association with the centre and periphery of the nucleus has been associated with transcriptional activation and silencing I was interested to see if there would be a difference in the nuclear localisation of 16p11.2 that was influenced by estrogen. I therefore analysed the FISH data to determine whether or not there were differences in the radial positioning of locus 2 under the different culture conditions.

The nuclear position of locus 2 was quantified from the relative proportion of hybridization signals across five shells of equal area eroded from the centre (shell 1) through to the periphery (shell 5) for each cell line and for each culture condition. This data showed that in MCF7 cells the largest proportion of hybridisation signals were centrally located (shell 1) but that this reduced dramatically upon hormone deprivation and could be rescued by the reintroduction of estrogen to the hormone deprived media (Figure 5.6). These changes in the distribution of locus 2 over shells 1-5 upon hormone deprivation and subsequent treatment with estrogen were all statistically significant (Fisher's test $p<0.05$, Table 5.3). This shows that there is a re-localisation of locus 2, representing the p11.2 domain of chromosome 16, that is mediated by estrogen whereby in its absence it moves away from the centre of the nucleus and in its presence moves towards the centre.

Table 5.3: The significance of changes in the distribution of hybridisation signals in MCF7 cells under different culture conditions(N – normal culture conditions, S-starvation conditions, E – estrogen-treatment). Fisher's exact test p-values

Cell Culture Conditions		Fisher exact test (p-value)
N	S	0.004
N	E	0.684
S	E	0.042

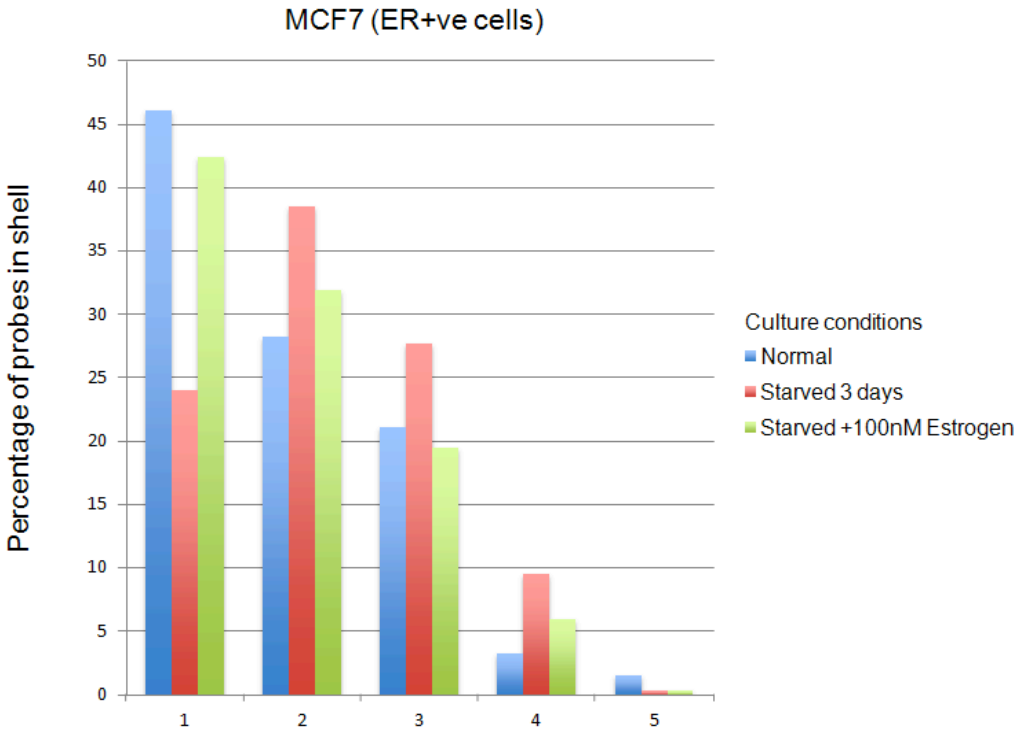


Figure 5.6: Nuclear erosion of MCF7 cells shows an ER-mediated re-localisation of locus 2. Shown are the percentage of FISH probe signals for this region that occurred in each of five shells of equal area with increasing shell number indicating shells away from the centre of the nucleus, the outermost (shell 5) representing the periphery.

Given the absence of any affect of estrogen withdrawal and re-addition on chromatin compaction in ER-ve MDAMB231 cells (Figure 5.5), I would not have expected to see a change in radial positioning of the locus in these cells either (Figure 5.7). There appeared to some subtle, but not statistically significant (Table 5.4) effect in these nuclei that are quite distinct from that seen in the ER+ve MCF7 cells

Table 5.4: The significance of changes in the distribution of hybridisation signals in MDAMB231 cells under different culture conditions (N – normal culture conditions, S-starvation conditions, E – estrogen-treatment). Fisher's exact test p-values

Cell Culture Conditions		Fisher exact test (p-value)
N	S	0.904
N	E	0.086
S	E	0.065

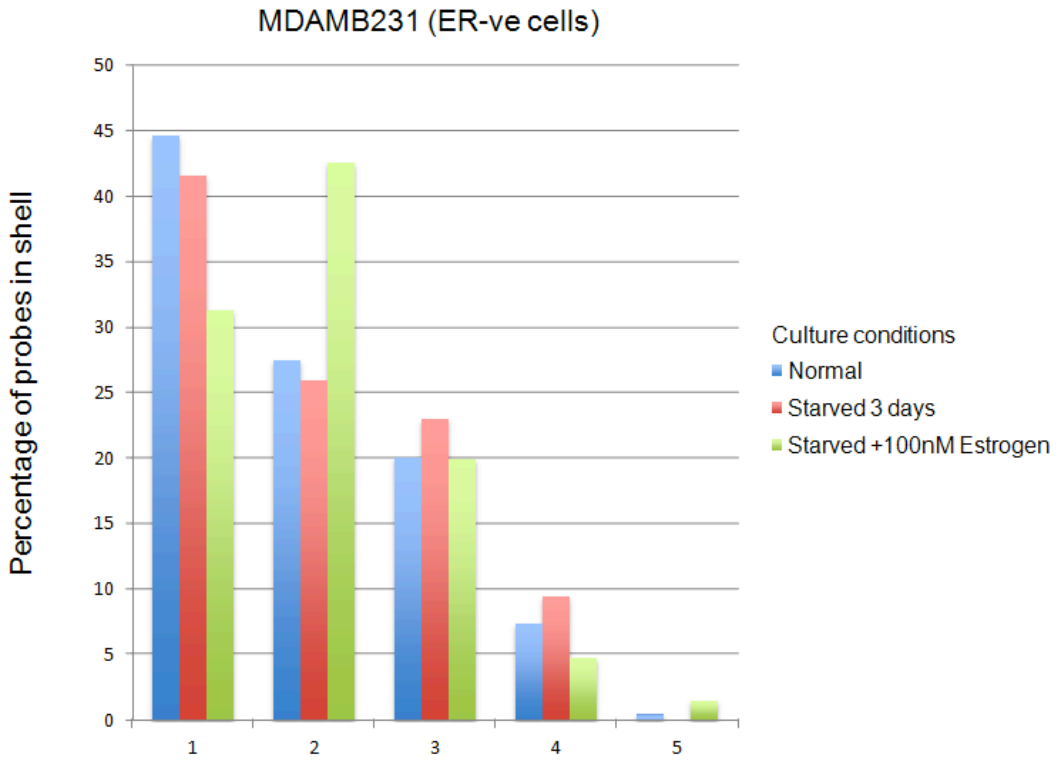


Figure 5.7: Nuclear erosion of MDAMB231 cells shows an ER-mediated re-localisation of locus 2. Shown are the percentage of FISH probe signals for this region that occurred in each of five shells of equal area with increasing shell number indicating shells away from the centre of the nucleus, the outermost (shell 5) representing the periphery.

5.2.3: Changes in nuclear area upon hormone deprivation and stimulation by estrogen

I also observed gross changes in the nuclear area of MCF7 and MDAMB231 cells under the different culture conditions (Figure 5.8 and Table 5.5). Although the nuclear area for the ER+ve cell line MCF7 was larger than the ER-ve MDAMB231 cells there were also differences when depriving both cell types of hormones and when reintroducing estrogen to the cells. Interestingly in both cases there was a significant reduction in the overall nuclear size (as assessed by DAPI staining) upon hormone deprivation and reintroduction of estrogen to the media by comparison to cells growing in their normal media. Moreover for MDAMB231 cells there was also a significant reduction in the nuclear size between hormone deprived cells and those that were hormone deprived and then treated with estrogen. This effect was not significant in the MCF7 cells. It is likely that these changes in nuclear area were due to cell cycle changes upon hormone deprivation and that adding estrogen back in the ER+ve MCF7 cells was sufficient to allow cell cycle progression (Prall et al., 1997).

Table 5.5: Summary table of all Wilcox test p-values for cell line analysis of nuclear area using FISH data

Cell lines tested		Wilcox test p-value
MCF7 (Normal)	MCF7 (Starved)	0.014
MCF7 (Normal)	MCF7 (Starved + Estrogen)	0.000
MCF7 (Starved)	MCF7 (Starved + Estrogen)	0.275
MDAMB231 (Normal)	MDAMB231 (Starved)	0.010
MDAMB231 (Normal)	MDAMB231 (Starved + Estrogen)	0.000
MDAMB231 (Starved)	MDAMB231 (Starved + Estrogen)	0.000

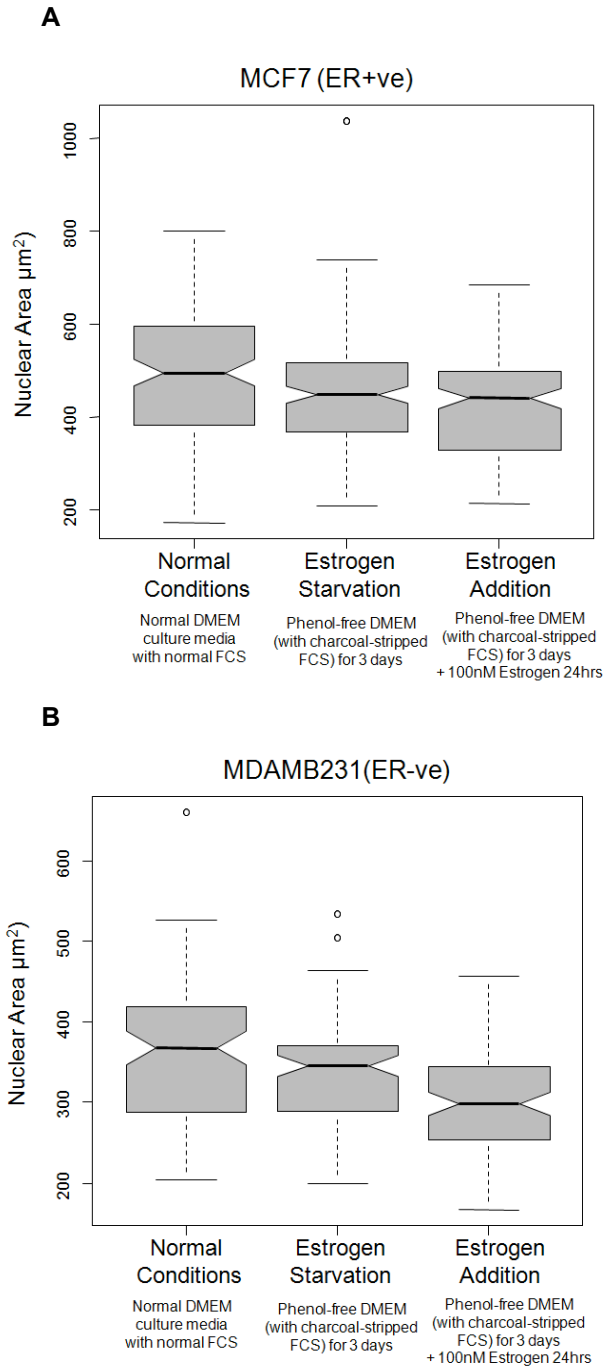


Figure 5.8: Nuclear area differences for MCF7 and MDAMB231 cells. The nuclear area (μm^2) distribution across all nuclei examined by FISH as determined by quantification of DAPI staining for (A) ER+ve MCF7 cells and (B) ER-ve MDAMB231 cells, under different culture conditions. Data from FISH analysis of locus 2 was used for this analysis.

5.3: Discussion

In the previous chapter I was able to pinpoint a sub-region of coordinated misexpression of genes on 16p11.2 that correlates with chromatin packing. There was also a marked difference in the localisation of the domain in the nucleus that was subtype specific and thought to place the RER in an environment that was conducive to active gene expression. The re-localisation of 16p11.2 towards the centre of the nucleus, coupled with its chromatin decompaction was highly correlated with the estrogen receptor status of the cells examined. This suggested that there are mechanisms in place at this locus that require the estrogen receptor to mediate the aberrant expression patterns both with regard to subtype and normal breast.

I investigated the role of the estrogen receptor with regard to the sub-RER that was aberrantly activated in ER+ve cell lines/tumours compared with ER-ve and normal cell lines/tissue. The estrogen receptor has been previously associated with large-scale changes in chromatin organisation through both extensive decondensation (Nye et al., 2002) and complex intrachromosomal interactions (Fullwood et al., 2009). I found that starving ER+ve cells of hormone resulted in a compaction of the locus that could be reversed by adding estrogen to the medium. This was not the case in the ER-ve cells tested. This effect was not seen at the control locus adjacent to the RER domain on chromosome 16p11.2. Furthermore these changes were associated with a re-localisation away from the nuclear centre in the absence of estrogen which was reversible in ER+ve cells by reintroducing estrogen to the system.

Chromatin decondensation and nuclear repositioning of loci that are being actively transcribed has been shown previously for the HoxB gene cluster in the developing embryo (Chambeyron et al., 2005). Therefore it is possible that in ER+ve cells the active gene expression signature is reliant on estrogen signalling to decompact locus 2 and reposition it to a hub of enriched in factors for active transcription. Certainly this region did not appear to be one of Fullwood et al's (2009) regions of estrogen induced complex looping domains to active transcription hubs nor Hsu et al's (2010) looped domains of transcriptional repression.

The nuclear area in the estrogen deprivation and treatment experiments was also grossly effected. There was an overall reduction in the size of the nucleus in both the ER+ve MCF7 cells and the ER-ve MDAMB231 cells after culturing in hormone deprived media. This would be expected given that the hormones are needed for the growth and propogation of cells in culture. This is in keeping with published data that inhibition of estrogen induces G1 phase arrest in the cell cycle and that subsequent treatment with estrogen allows re-entry of these cells into S phase after 12 hours (Prall et al., 1997). This would help explain why the nuclear area was not significantly reduced in ER+ve MCF7 cells after estrogen addition (24 hours) compared to ER-ve cells, if indeed this is a cell cycle effect. This could be tested by Fluorescence Activated Cell Sorting (FACS) analysis of cell populations under each of the culture conditions tested to see if there is a change in the proportion of cells in different stages of the cell cycle.

Chapter 6

Comparing RER domains in breast cancer to those in hepatoblastoma and published LRES domains

6.1: Introduction

In Chapter 3 I identified domains of copy-number-independent coordinate misregulation of gene expression in breast cancer. I went on to show that the aberrant activation of one of these domains on chromosome 16p11.2 was correlated with an overall decompaction of the chromatin at this locus and that this effect was tumour subtype specific (Chapter 4). Moreover the changes in chromatin compaction between ER+ve and ER-ve cell lines was estrogen dependent (Chapter 5).

During my PhD, as part of a collaboration with Marie Annick Buendia (INSERM U785, Centre Hepato-Biliaire, Hôpital Paul Brousse, Villejuif, France) who was interested in investigating epigenetic changes in hepatoblastoma I was asked to identify RER domains in this tumour subtype. I was able to use the methods described in Chapter 3 to identify RER domains in hepatoblastoma tissues using published expression and CGH array data from the Buendia lab. This chapter briefly summarises the findings from this collaboration.

To identify whether there were "universal" RER domains that were not cancer type specific, I have cross referenced my findings for regions of epigenetic regulation in breast cancer with those I derived for hepatoblastoma. To this end I also revisit the LRES domains described in published cancer studies and compare them with the breast cancer RERs.

6.2: Results

6.2.1. RER in Hepatoblastoma

Hepatoblastomas are childhood cancers that are thought to originate from liver precursor cells. Wnt signaling and c-myc function are thought to be important in this malignancy (Cairo et al., 2012). Using aCGH and expression data for 25 liver hepatoblastomas and 4 normal tissues I used the computational approach described in chapter 3 to identify RER domains in the liver hepatoblastoma samples (Cairo et al., 2008).

Overall, 3.9% of the genes showed a significant correlation in expression (TCS) with their neighbours that was copy-number independent. DNA copy number profiles for the 25 tumour samples and TCS maps for all chromosomes are shown in Appendix XI with examples in Figure 6.1. The genes with significant TCS were grouped into 68 regions with more than 1 significant TCS gene and containing overall 329 genes above the threshold set by transforming the data into z-scores and setting a significance cut off of $p<0.05$ (as described previously for breast cancer). A full list of the 68 RER domains with TCS genes in the regions is shown in Table 6.1.

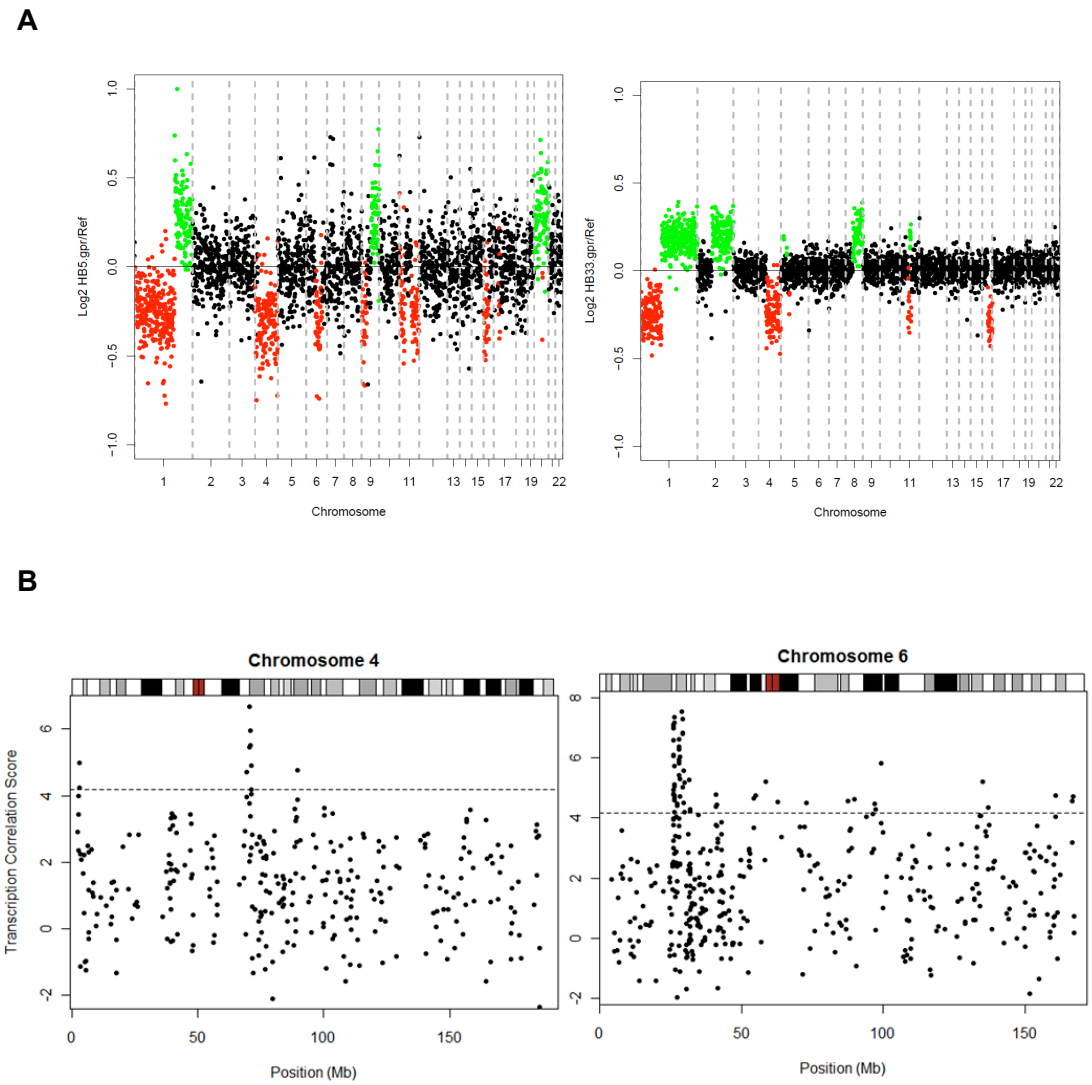


Figure 6.1: Copy number profiles & Transcription Correlation Maps (TCM's) in Hepatoblastoma. (A) The log2 (sample:ref) ratios of the clones were plotted against chromosomal position. Regions of loss (green), gain (red), and no change (black) as determined by CGHcall (van de Wiel et al., 2007) (B) Transcriptome Correlation Maps generated for chromosomes 4 & 6 using data for 25 hepatocoma samples. Dashed line represents the significance threshold. Scores above this threshold are regions of coordinated expression.

Mean centred z scores were calculated for all genes in the region and a comparison of overall gene expression in the RERs to that of normal tissue showed that hepatoblastoma, like breast cancer, does not conform to the simple LRES phenotype of misregulation of large genomic domains. Overall I found that of the 68 RER domains, just 17 of them were significantly downregulated in tumours relative to normal tissue and 14 were significantly upregulated. Moreover there were 37 RER domains where the overall gene expression for the region was not significantly different between normal and tumour tissue. Examples of aberrantly up/down-regulated regions are shown in Figure 6.2 and highlighted red/green respectively in Table 6.1 with Wilcox p-values shown for all 68 regions.

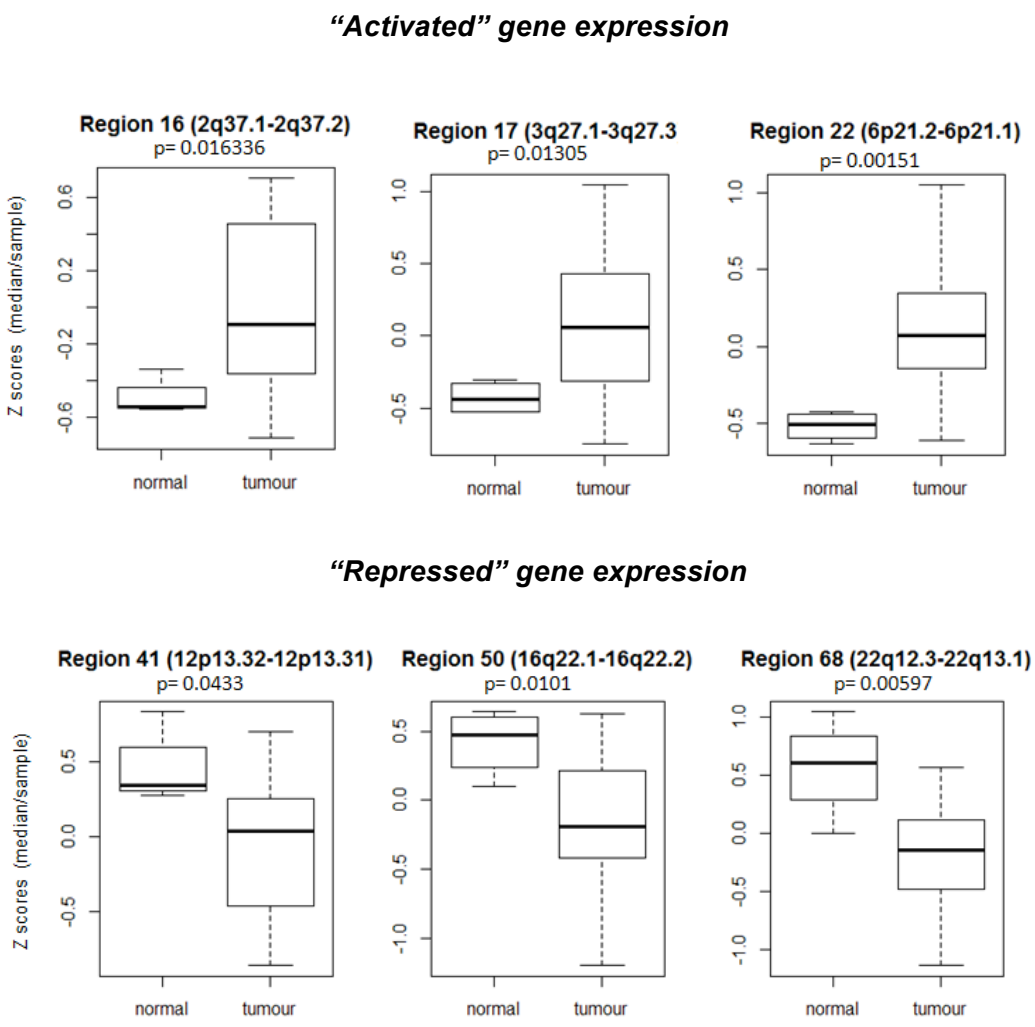


Figure 6.2: Analysis of gene expression changes in tumours relative to normal tissue. Boxplots showing the distribution of median z score between samples. (A) Regions 16, 17 and 22 - three examples of regions that were relatively upregulated in tumours. (B) regions 41, 50 and 68, are three examples of regions that were relatively downregulated in tumours.

Table 6.1. Summary of 68 RER domains in hepatoblastoma. Significant up/down-regulated RERs were obtained by comparing tumour expression to normal liver tissue samples (Wilcox p<0.05). RERs highlighted in yellow show overlap with the 26 RERs identified from breast cancer cell lines and tumours.

RER Identifier	Chromosome band	Size (Mb)	Significant TCS genes	Hepatoblastoma V Normal Liver
1	1p36.23 - 1p36.21	7.97	*PGD,DFFA,TNFRSF8,PRAMEF10*	No.sig.change
2	1p36.13 - 1p36.12	3.44	*PAX7,RNF186*	No.sig.change
3	1p34.2 - 1p34.1	2.69	*CDC20,DPH2,B4GALT2,RNF220, TMEM53*	Downregulated
4	1p32.3 - 1p32.1	7.24	*DIO1,HSPB11,DHCR24,GOT2L1, PRKAA2*	No.sig.change
5	1p12 - 1q21.2	27.81	*ADAM30,ITGA10,CD160*	No.sig.change
6	1q21.3	1.65	*CRNN,C1orf46*	No.sig.change
7	1q21.3 - 1q22	1.27	*RAG1AP1,SCAMP3,CLK2,YY1AP1,DAP3*	Downregulated
8	1q23.1 - 1q23.2	2.57	*CD5L,KIRREL,CD1D,CD1A,CD1C*	No.sig.change
9	1q24.2 - 1q25.1	4.66	*FMO4,C1orf105*	No.sig.change
10	1q25.1 - 1q25.3	6.02	*RALGPS2,ABL2*	Downregulated
11	1q32.1 - 1q32.3	8.52	*MDM4,LRRN2,CNTN2,AVPR1B,DYRK3, IL19,IL24,C4BPB,C4BPA,CR2,HSD11B1,IRF6*	Upregulated
12	1q42.12 - 1q42.3	11.76	*H3F3A,JMJD4,HIST3H2A,NUP133,URB2,PGBD5, COG2,AGT*	Downregulated
13	2p23.3 - 2p23.2	4.26	*DTNB,KCNK3,CAD,PPM1G*	No.sig.change
14	2q31.1 - 2q32.2	13.75	*UBE2E3,SSFA2*	No.sig.change
15	2q35 - 2q36.3	8.38	*IHH,NHEJ1,ABCB6,ANKZF1,TUBA4A, SPEG,GMPPA,SLC4A3*	No.sig.change
16	2q37.1 - 2q37.2	3.32	*NEU2,SAG*	Upregulated
17	3q27.1 - 3q27.3	2.53	*PSMD2,POLR2H*	Upregulated
18	4p16.3 - 4p16.2	3.27	*TETRAN,NOP14*	Upregulated
19	4q12 - 4q13.3	14.99	*TMPRSS11E2,UGT2B4,SULT1B1, CSN1S1,STATH,HTN3,CSN3,SMR3B*	Upregulated
20	6p22.2 - 6p21.33	5.45	*HIST1H1A,HIST1H4A,HIST1H2AB, HIST1H2BB, HIST1H3C,HIST1H1C, HIST1H1T,HIST1H2BC, HIST1H2AC, HIST1H1E,HIST1H4D,HIST1H4E, HIST1H2AE,HIST1H3E,HIST1H1D, HIST1H4F, HIST1H2APS4, RP1-5P21.3,POM121L2, ZNF391,HIST1H2BL, HIST1H2BM,HIST1H2BN, HIST1H2AL, HIST1H1B, HIST1H2BO,OR2B2, ZNF192, ZKSCAN3,ZSCAN12,OR2W1, OR2J3,OR2N1P,OR12D3, OR12D2, OR11A1, OR2H1,HCG9,RNF39*	No.sig.change
21	6p21.33	1.04	*NCR3,C6orf25,C6orf27*	Upregulated
22	6p21.2 - 6p21.1	4.08	*APOBEC2,TREM2,TREML2*	Upregulated
23	6p12.2 - 6q13	19.73	*RP3-334F4.2,HCRTR2, RP11-278J20.1, KHDRBS2*	Upregulated
24	6q14.2 - 6q21	22.21	*HTR1E,GABRR2,FHL5,GPR63,POU3F2*	No.sig.change
25	6q23.2 - 6q24.1	9.41	*ALDH8A1,MAP3K5*	No.sig.change
26	6q25.3 - 6q27	7.41	*RP11-72O9.4, T,RP3-497J21.1*	Upregulated
27	7q21.3 - 7q22.1	4.85	*SLC25A13,ACN9*	Upregulated
28	7q34 - 7q36.1	10.97	*TRBV10-2,TRBV5-2,TRBV21-1, RPV5,CLCN1 ,FAM131B,ZNF767, ZNF862,RARRES2,REPIN1, GIMAP4*	No.sig.change
29	8q11.21 - 8q12.1	7.30	*TCEA1,LYPLA1*	No.sig.change
30	8q22.1 - 8q22.3	5.55	*COX6C,ZNF706*	No.sig.change

31	8q24.22 - 8q24.3	12.00	*AC115836.1,LY6D,ZNF696,PUF60,CYC1,DGAT1*	No.sig.change
32	9p22.1 - 9p21.1	13.48	*IFNB1,IFNA5,IFNA2,IFNA8,C9orf53*	Upregulated
33	9q33.3 - 9q34.11	1.89	*DPM2,PTGES2,COQ4,URM1,ODF2*	No.sig.change
34	10q11.22 - 10q21.3	17.76	*SLC18A3,A1CF,CISD1*	Downregulated
35	10q22.3 - 10q23.31	9.12	*LRIT1,RGR,MMRN2*	Downregulated
36	10q24.31 - 10q24.32	2.13	*LBX1,POLL,FGF8*	No.sig.change
37	11p15.4	2.82	*OR52A1, HBE1*	Upregulated
38	11p15.1 - 11p13	13.22	*LGR4,KCNA4,ELP4,PAX6*	Upregulated
39	11q12.2 - 11q12.3	1.36	*SCGB1D1,SCGB2A1,SCGB1A1,MTA2*	Upregulated
40	11q13.1 - 11q13.2	0.57	*BANF1,SF3B2,RAB1B*	No.sig.change
41	12p13.32 - 12p13.31	2.75	*FGF23,FGF6,AKAP3,GALNT8*	Downregulated
42	12p13.31 - 12p13.2	2.75	*KLRB1,CD69,KLRF1,CLEC2B,CLEC7A, KLRK1*	No.sig.change
43	12q13.11 - 12q13.12	2.14	*WNT10B, TROAP*	No.sig.change
44	12q13.13	1.97	*NR4A1,KRT85,KRT84,KRT75,KRT6B, KRT5,KRT2,KRT1,KRT76*	Downregulated
45	12q13.13	0.92	*HOXC13,HOXC11*	No.sig.change
46	14q13.3 - 14q21.3	13.76	*FKBP3,RPS29*	No.sig.change
47	14q32.33	0.86	*IGHA1,IGHG1*	No.sig.change
48	15q15.3 - 15q21.2	6.53	*SEMA6D,SLC12A1*	Downregulated
49	16p13.3	1.55	*DECR2,RAB40C,RHOT2,TPSG1,TPSD1*	No.sig.change
50	16q22.1 - 16q22.2	3.13	*TERF2,AARS,ST3GAL2*	Downregulated
51	17p13.3 - 17p13.2	1.86	*OR1D2,AC097370.1,OR1G1,OR1A2, OR1A1,OR3A2,OR3A1,OR1E1, OR3A3,OR1E2*	No.sig.change
52	17q11.2	1.50	*TMEM199,SARM1,SLC13A2,FOXN1, UNC119*	No.sig.change
53	17q21.2	2.08	*KRTAP1-3,KRTAP1-1,KRTAP2-4, KRT33B,KRT34,KRT31,KRT32,KRT35, KRT36,KRT13,KRT9,KRT16,KLHL11*	No.sig.change
54	17q21.31	2.96	*RUNDC3A,GJC1,C1QL1*	Downregulated
55	18q21.31 - 18q22.1	8.28	*TNFRSF11A,ZCCHC2,SERPINB4, SERPINB7,CDH7*	No.sig.change
56	19p13.3	3.19	*ATP5D,RPS15,MBD3,SCAMP4,BTBD2, MKNK2,PLEKHJ1,SGTA,TLE6,GNA15, DOHH*	No.sig.change
57	19p13.3	1.75	*FUT5,RFX2,ACSBG2*	Downregulated
58	19p13.2	1.35	*CDC37,DNM2*	No.sig.change
59	19p13.12 - 19p13.11	2.04	*OR7C2,SLC1A6,OR10H2*	No.sig.change
60	19p13.11	1.89	*RAB3A,PGPEP1,TMEM59L,COPE, ARMC6*	No.sig.change
61	19q13.12 - 19q13.32	8.05	*CD79A,POU2F2,MEGF8,PSG9,ZNF428,ZNF223, ZNF224*	Downregulated
62	19q13.33 - 19q13.41	1.87	*SHANK1,GPR32,KLK1,KLK3, KLK5,KLK6, KLK7,KLK8,KLK11,KLK14,SIGLEC9*	No.sig.change
63	19q13.42	1.41	*AC006293.3,NCR1,TNNT1,SYT5*	Downregulated
64	19q13.43	0.72	*ZNF606,ZNF329*	No.sig.change
65	20q13.33	1.47	*COL9A3,C20orf11,ARFGAP1,EEF1A2, PTK6*	Downregulated
66	22q11.21	0.95	*SLC25A1,CLTCL1*	No.sig.change
67	22q11.21 - 22q11.23	2.65	*IGLV6-57,PRAME*	Downregulated
68	22q12.3 - 22q13.1	1.55	*KCTD17,SH3BP1*	Downregulated

To understand the function of the genes occurring in RER domains, I performed Gene Ontology (GO) analysis using the GO web-based program Gorilla (Eden et al., 2009, 2007). GO term analysis by biological function showed enrichment for receptor activity in the RER domains and most significantly olfactory receptor activity (summarised in Table 6.2).

The olfactory receptor (OR) genes comprise a large family of genes organised in clusters and distributed across all but a few chromosomes (Rouquier et al., 1998). As well as being expressed in sensory neurons they are also expressed in other tissues where their function is largely unknown (Flegel et al., 2013). Their altered expression has been reported in both prostate and breast cancer where they impact tumour progression (Muranen et al., 2011; Neuhaus et al., 2009).

Ephirin receptor activity was also highly enriched by GO term analysis which was not surprising given that as a member of the receptor tyrosine kinase family it has been associated with a wide range of tumour types with multi-faceted roles in tumorigenesis.

GO term analysis by biological process showed significant enrichment for epidermal and tissue development as well as factors that affect chromatin organisation like nucleosome assembly and organization (Table 6.3).

Table 6.2. GO term analysis by biological function

GO Term	Description	P-value	FDR q-value	Enrichment
GO:0004984	olfactory receptor activity	1.26E-09	3.84E-06	3.72
GO:0005198	structural molecule activity	4.60E-05	7.00E-02	1.5
GO:0004872	receptor activity	8.20E-04	8.32E-01	1.28
GO:0004888	transmembrane signaling receptor activity	8.22E-04	6.26E-01	1.33
GO:0005003	ephrin receptor activity	8.66E-04	5.27E-01	3.99

Table 6.3. GO term analysis by biological process

GO Term	Description	P-value	FDR q-value	Enrichment
GO:0008544	epidermis development	2.18E-08	2.04E-04	2.55
GO:0006334	nucleosome assembly	3.84E-06	1.80E-02	2.5
GO:0034728	nucleosome organization	2.39E-05	7.47E-02	2.25
GO:0009888	tissue development	2.87E-05	6.73E-02	1.54
GO:0065004	protein-DNA complex assembly	3.93E-05	7.37E-02	2.24
GO:0071824	protein-DNA complex subunit organization	8.87E-05	1.39E-01	2.11
GO:0050907	detection of chemical stimulus involved in sensory perception	1.89E-04	2.53E-01	2.76
GO:0006705	mineralocorticoid biosynthetic process	2.34E-04	2.74E-01	5.32
GO:0008212	mineralocorticoid metabolic process	2.34E-04	2.44E-01	5.32
GO:0006575	cellular modified amino acid metabolic process	5.69E-04	5.34E-01	1.69
GO:0031424	keratinization	8.02E-04	6.84E-01	3.27
GO:0006704	glucocorticoid biosynthetic process	8.66E-04	6.77E-01	3.99

6.2.2. Comparing Breast Cancer & Hepatoblastoma RER's

The hepatoblastoma RER domains were compared to those identified for breast cancer and are visualised as a circular map in Figure 6.3. Although ten of RER domains (shaded yellow in Table 6.1) in hepatoblastoma overlapped all/in part with the 26 RER's I found for the breast cancer cell line and tumour tissue data, only two of these regions showed a significant (Wilcox test $p<0.05$) aberrant change in gene expression when compared to normal liver tissue. These were the 3.13 Mb RER on chromosome 16q22.1-22.1 and the 2.64 Mb RER on chromosome 22q11.21-q11.23. These two domains were both aberrantly repressed in hepatoblastoma and breast tumours (Table 6.1).

Figure 6.3: Circular map of RER domains identified for breast and liver cancer. Shown are the RER domains on each chromosome derived from analysis of aCGH and expression array data for breast (outer circle, red) and liver (inner circle, blue) tumour data. RER domains for breast are based upon the 26 overlapping domains obtained from comparison of cell line and tissue RERs.

6.2.3. Comparison of RER domains to LRES domains in published studies

I also compared the RER data for breast cancer with other published studies that have set out to systematically identify LRES domains in cancer. Hsu et al (2010) identified 11 domains of LRES that were estrogen-mediated in breast cancer however only one of these domains was a significant RER domain from my analysis of breast cancer cell lines and tumours. This was the RER on chromosome 16p11.2 that contained the 14 gene region falling into the RER domain (locus 1) that I investigated during my PhD. Given that Hsu et al were looking specifically for domains silenced in response to estrogen signalling and thereby used different criteria and methodology to mine, this finding is not surprising.

Comparison of the 26 overlapping RER domains common to both my breast tumour tissue and cell line analysis with the 28 LRES regions in bladder carcinoma described by Stransky et al (2006) showed no common regions. However on comparison of the 45 breast tumour and 71 breast cancer cell line RER's showed 10 RER domains that appear as LRES domains in bladder carcinoma (Table 6.2). This included the 3-2 domain on 3p22.3 that was followed up in that study and found to be associated with increased histone methylation (H3K9me3 and H3K27me3) and histone hypoacetylation, without any associated with DNA methylation changes (Stransky et al., 2006).

Table 6.2: Summary of comparison between LRES domains in bladder carcinoma with RER domains in breast cancer cell lines and tumours

Bladder TCS region (identification number)	Cytoband	Significant TCS genes in bladder carcinoma (after recalculation)	Size (bp) in bladder carcinoma	Breast Tumour RER (identification number)	Breast Cell line RER (identification number)
4-2	4q13.3	*IL8, CXCL6, CXCL1, CXCL2, CXCL5, CXCL3*	296	-	11
6-7	6q23.3–6q24.1	*MAP7, MAP3K5, PEX7, IFNGR1, HEBP2, C6orf80, CITED2*	3029	7	-
14-1	14q11.2	*LRP10, ACIN1, PABPN1, EFS, AP1G2, DHRS2,PCK2, WDR23, PSME1*	1,260	24	-
19-2	19p13.13	*PRDX2, FARSLA*	125	-	57
19-3	19p13.12–19p13.11	*WIZ, CYP4F3, CYP4F12, CYP4F11, CHERP, SIN3B, MYO9B,FLJ22709, NR2F6, FCHO1, B3GNT3, INSL3, SLC5A5, PIK3R2*	2,732	-	58
6-5	6q16.1	*KIAA0776, C6orf111, ASCC3, PREP*	8,755	6	-
1-4	1p34.1	*ATP6V0B, PRNPIP, (RPS15A), (UROD), (PRDX1), (AKR1A1)*	2,108	-	3
17-8	17q21.33–17q22	*WDR50, COX11*	3,701	36	-
3-2	3p22.3	*PLCD1, ACAA1, MYD88*	277	-	8
3-5	3p21.31	*IMPDH2, GPX1, RHOA, AMT, UBE1L, MST1R*	1,288	3	-

Comparison of my 26 RER domains with the 47 LRES regions in prostate cancer described by Coolen et al (2010) showed two common regions: one 3.2 Mb domain on chromosome 8q22.3-23.1 and the other a 1.06 Mb domain on chromosome 10q26.13 (Table 6.3, highlighted yellow). Further comparison of the 45 breast tumour and 71 breast cancer cell line RER's showed a further 11 RER domains that appeared as LRES domains in prostate cancer (Table 6.3). Notably 10 of these 11 RER's were breast tumour data derived.

Table 6.3: Summary of comparison between LRES domains in prostate cancer with RER domains in breast cancer cell lines and tumours

Prostate LRES (identification number)	Cytoband	LRES genes within region	Size (Mb) in prostate cancer	Breast Tumour RER (identification number)	Breast Cell line RER (identification number)
8	2q14.2–q14.3	*RALB, INHBB, GLI2, TFCP2L1, RNU4ATAC, CLASP1*	1.4	2	-
11	3p14.3	*SLMAP, (FLNB), (DNASE1L3), (ABHD6), (RPP14), (PXK), (PDHB), (KCTD6), (ACOX2), (FAM107A)*	0.83	3	-
18	6p12.1–p11.2	*(HCRTR2), (GFRAL), (HMGCLL1), (BMP5), (COL21A1), (DST), (BEND6), (KIAA1506), (ZNF451), BAG2, RAB23, PRIM2*	2.48	6	-
19	6q14.3–q15	*CYB5R4, MRAP2, KIAA1009, TBX18, NT5E, SNX14, SYNCRIP, SNORD50A, SNORD50B, SNHG5, HTR1E, CGA, ZNF292, GJB7, C6orf162, C6orf165, SLC35A1, RARS2, ORC3L, NCRNA00120, AKIRIN2, SPACA1, CNR1, RNGTT*	5.11	6	-
20	6q21	*PDSS2, SOBP, SCML4, SEC63, OSTM1, NR2E1, SNX3*	1.11	6	-
26	8q22.3–q23.1	*DPYS, LRP12, ZFPM2, OXR1, ABRA, ANGPT1*	3.2	12	26
31	10q26.13	*FGFR2, ATE1, NSMCE4A, TACC2, BTBD16, PLEKHA1, ARMS2, HTRA1*	1.06	17	32
34	12q21.2	*ZDHHC17, CSRP2, E2F7, NAV3, SYT1, PAWR, PPP1R12A*	3.17	20	-
36	16q12.2–q13	*(MMP2), (LPCAT2), (CAPNS2), (SLC6A2), (CES4), (CES1), (CES7), (GNAO1), AMFR, NUDT21, OGFOD1, BBS2, MT4, MT3, MT2A, MT1L, MT1E, MT1M, MT1A, MT1DP, MT1B, MT1F, MT1G, MT1H, MT1IP, MT1X*	1.21	30	-
37	16q23.3–q24.1	*MBTPS1, HSDL1, LRRC50, TAF1C, ADAD2, KCNG4, WFDC1, ATP2C2, KIAA1609, COTL1, KLHL36, (USP10), (CRISPLD2), (ZDHHC7), (KIAA0513)*	1.04	30	-
39	18p11.22–p11.21	*RAB31, TXNDC2, VAPA,APCDD1, NAPG, FAM38B, GNAL, CHMP1B, MPPE1*	2.20	39	-
41	22q12.3	*C22orf28, BPIL2, FBXO7, TIMP3, SYN3, LARGE*	1.53	43	-
42	22q13.1	*APOBEC3C, APOBEC3D, APOBEC3F, APOBEC3G, APOBEC3H, CBX7, PDGFB, SNORD83B, SNORD83A, RNU86, SNORD43, RPL3, SYNGR1*	0.37	-	71

One of the first LRES domains identified was the HoxA gene cluster on chromosome 7p15.2 (Novak et al., 2006). Subsequently this LRES domain was also identified as a differentially methylated region (DMR) in breast cancer (Novak et al., 2008). Therefore I would have expected the HoxA gene cluster to appear as an RER domain in my analysis, but this was not the case. Further analysis showed that in my breast cancer cell line analysis a number of genes at this locus had high TCS but they did not meet the threshold of significance to denote it a RER domain (Figure 6.4). The datasets I had for breast tumour analysis however did not include many of the genes at this locus and those that were present did not have high TCSs. Upon investigation of the other DMR domains described in (Novak et al., 2008) I found that the RER containing the histone gene cluster was also a DMR hypermethylated in breast cancer (Table 6.4). However, in my analysis, the genes in this cluster were not significantly up or down-regulated relative to normal tissue. Co-ordinate regulation of histone gene clusters is to be expected given that they are generally replication-dependent reaching high levels during rapid S phase (Marzluff et al., 2008). I also found one hypermethylated and one hypomethylated DMR that corresponded with cell line RER domains identified by my analysis (Table 6.4).

Table 6.4: Summary of comparison between DMR domains in prostate cancer with RER domains in breast cancer cell lines and tumours

DMR status	Cytoband	Genes within DMR that overlap	DMR size (Kb)	Breast Tumour RER (identification number)	Breast Cell line RER (identification number)
Hypermethylated	6p22.2	*HIST1H2BE, HIST1H4D, HIST1H3D, HIST1H2AD, BC056264, HIST1H2BF, HIST1H4E, BC082232, HIST1H2BG, HIST1H2AE, HIST1H3E, HIST1H1D, HIST1H4F, HIST1H4G, HIST1H3F, HIST1H2BH, HIST1H3G, HIST1H2BI*	91	5	15
Hypermethylated	6p21.32	*RXRB, X66424, SLC39A7, HSD17B8, RING1*	260	-	17
Hypomethylated	19q13.42	*KIR2DL1, KIR2DL3, KIR2DL4, KIR2DS2, KIR2DS4, KIR3DL1, KIR3DL2, KIR3DL3, KIR3DP1, L76664, L76668, LILRB4*	213	-	61

The protocadherin family cluster (PCDH) has been reported to be subject to LRES by hypermethylation in a number of tumour types including breast (Dallosso et al., 2012, 2009; Novak et al., 2008). I therefore investigated this cluster of genes on chromosome 5q31.3 in my transcriptional correlation data. As with the HoxA gene cluster I found that in both the cell line and tumour data the TCS genes did not meet my significance threshold (Figure 6.5). Also, as before, the tumour data contained very few of the protocadherin cluster genes and those that were present did not have high TCS values.

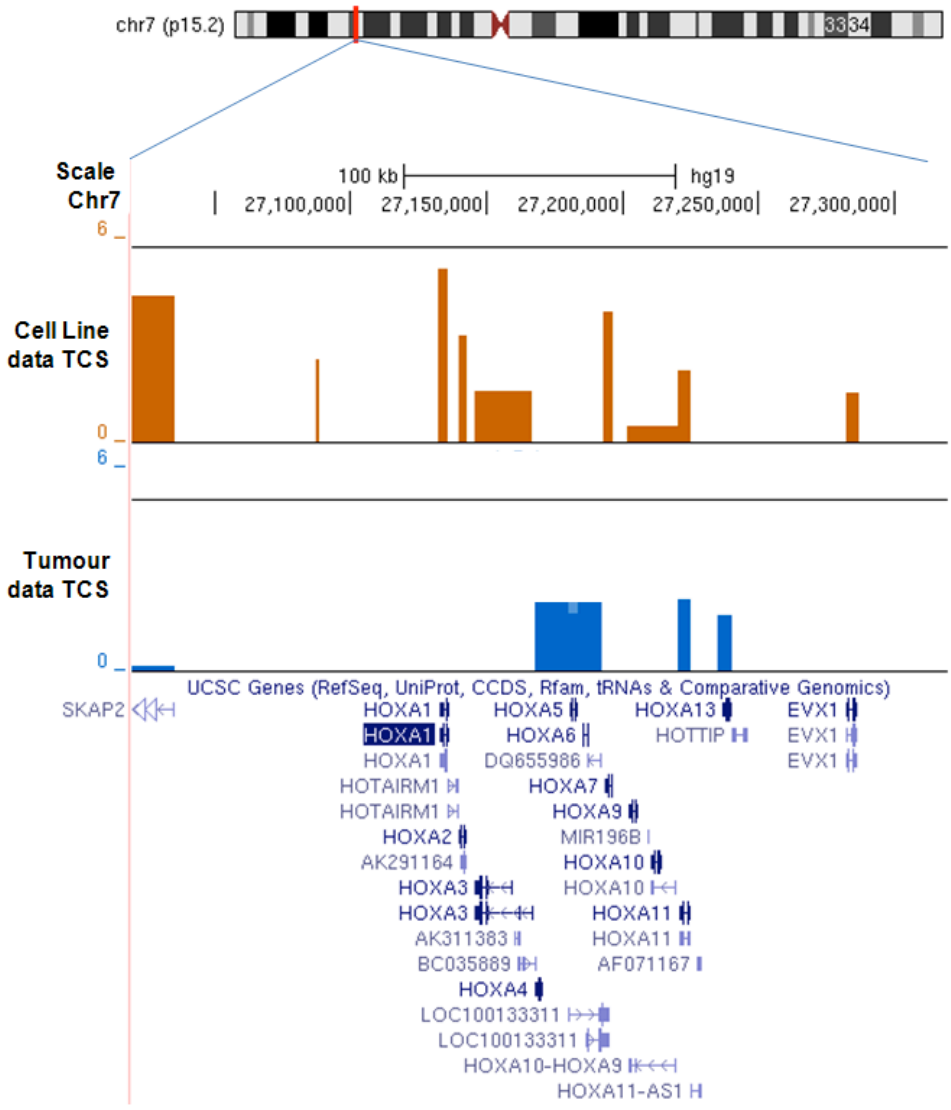


Figure 6.4: Transcription Correlation Scores (TCS) at the locus containing the HoxA gene cluster on chr7p15.2. Customised UCSC browser tracks with TCS in bedgraph format for breast cancer cell line (orange) and tumour data (blue). The horizontal line represented the TCS significance threshold. UCSC genes in their chromosomal position are shown at the bottom.

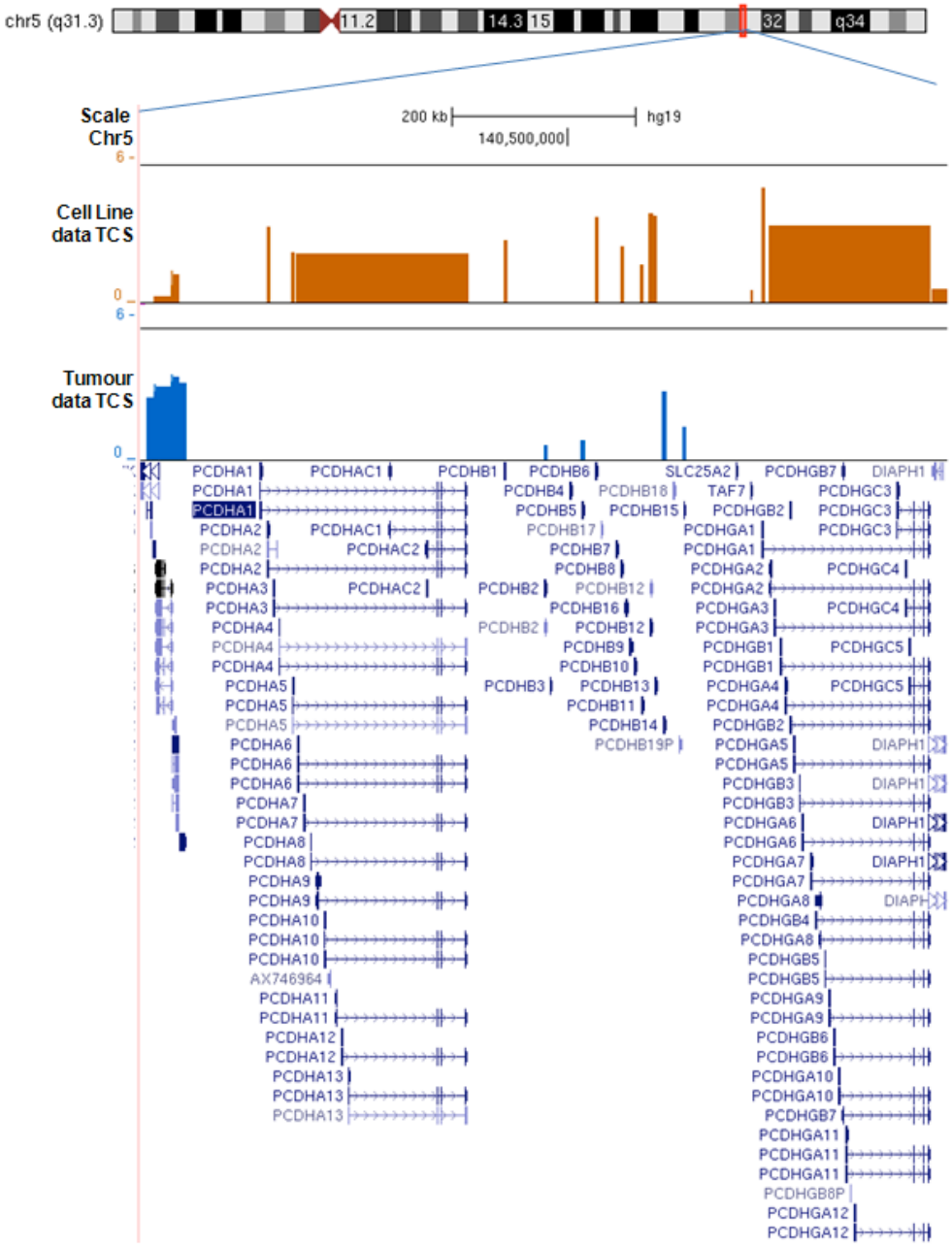


Figure 6.5: Transcription Correlation Scores (TCS) at the locus containing the protocadherin gene cluster on chr5q31.3. Customised UCSC browser tracks with TCS in bedgraph format for breast cancer cell line (orange) and tumour data (blue) Horizontal line represents the TCS significance threshold. UCSC genes in their chromosomal positon are shown at the bottom.

6.3: Discussion

Previously in this thesis I devised a method to identify regions of coordinate misregulation of gene expression that were copy number independent. This was to find regions that were potentially epigenetically regulated (RER domains) through changes in higher order chromatin architecture. Having validated this approach in breast cancer cell lines and tumours, it was possible to use this system to identify RER domains in other types of cancer or disease for which copy number and expression data are available.

In this chapter I described work carried out as collaboration with Marie Annick Buendia for hepatoblastoma. From this analysis I found a large number of genomic regions that were subject to RER. Interestingly the RER domains were very different to those I identified for breast cancer and where there was commonality those regions showed no significant change in gene expression between normal and tumour tissues. The two common regions between breast and liver that showed a significant difference in gene expression between normal and tumour tissue were both aberrantly downregulated. This concurs with the most frequently reported type of region of epigenetic regulation (RER) in the literature to date – the long range epigenetically silenced (LRES) phenotype.

GO term analysis of hepatoblastoma RER genes alluded significantly to both receptor activity and nucleosome organisation/assembly. This was particularly interesting as there was high enrichment for olfactory receptor (OR) activity and though these genes are abundantly expressed in a number of tissues little is known about their function (Flegel et al., 2013; Muranen et al., 2011; Neuhaus et al., 2009). Moreover chromatin organisation of the OR genes has been shown to govern their expression in olfactory neurons whereby they form heterochromatic foci which they can be actively transcribed outside of (Clowney et al., 2012). Therefore they also appear to have a role to play in nuclear organisation of genetic material which is suggestive of the mechanism of aberrant misregulation for these RER domains.

The main focus of my PhD was to investigate RER in breast cancer. Overall I found very few RER domains that corresponded to liver hepatoblastoma RER domains. In comparison to other types of cancer in published literature (i.e. prostate and bladder cancer) where similar computational methods have been used to systematically identify LRES, I found that few of these studies showed overlap with the RER domains from my analysis. This indicates that most of RER is cancer-specific in nature though there are some regions which overlap. Investigation of breast cancer studies also showed that some but not all of reported LRES regions were RER domains also and upon investigation this was mainly due to a depletion of those genes in the data used for analysis or that they missed the significance threshold.

Chapter 7

General discussion

Misregulation of gene expression is a common event in cancer with a number of long-range events documented. Such studies have uncovered large chromosomal domains of repression in association with a cocktail of DNA methylation changes, gain of repressive histone marks and loss of active histone marks. This phenomena was denoted as Long Range Epigenetic Silencing (LRES) and these studies are described in more detail in the Introduction (Section 1.3.5) to this thesis. At the start of my PhD the most interesting aspect of these LRES studies, in my opinion, was that (1) there does not seem to be any consensus as to the cause of misregulation of such large genomic regions with a number of lower level epigenetic changes reported and that (2) these studies all focussed on a deregulation of large domains by epigenetic repressive mechanism(s). My main hypothesis was that for such large regions of the genome to be co-ordinately misregulated there could be higher order epigenetic changes occurring involving the chromatin architecture itself that are having an effect on gene expression in these regions. Therefore the main aim of this project was to identify regions of potential epigenetic regulation in the breast cancer genome, which could be used to investigate this hypothesis and provide insight into the mechanism of LRES.

The main findings of my PhD have been that:

(1) In breast cancer there are large domains containing clusters of genes that are co-ordinately regulated.
(2) These domains show coordinate expression patterns that are independent of copy number aberrations, which could otherwise have accounted for these patterns.
(3) These RER domains can be found to overlap in breast tumour tissue and cell line datasets but there are also a number of regions that are tumour tissue or cell line specific.
(4) As demonstrated for breast cancer and liver hepatoblastoma, there are RERs that are subject to long range epigenetic activation (LREA), not just the LRES phenotype, and others that showed no significant change in gene expression when compared to normal tissue.

(5) Some of the breast cancer RERs showed gene expression signatures that were subtype-specific and were able to distinguish luminal/ER+ve from basal/ER-ve samples.
(6) The subtype-specific gene expression patterns correlated with changes in chromatin architecture and organisation.
(7) The changes in gene expression between normal breast tissue and tumour tissue correlated with changes in chromatin architecture and organisation
(8) The changes in chromatin architecture and organisation included: different levels of chromatin compaction and the nuclear localisation of the RER on 16p11.2
(9) The changes in chromatin architecture/organisation are estrogen-dependent
(10) Long range changes in gene expression by RER are mostly cancer specific with little overlap between cancer types

7.1: Computational biology as a tool for identifying RER

As discussed in section 1.1.2.5 a number of large scale molecular profiling studies using transcriptome data has allowed the classification of breast cancers into distinct groups based on their gene signature which is predictive of clinical outcome (Blows et al., 2010; Lu et al., 2012; Prat et al., 2010; Sørlie et al., 2001; Sorlie et al., 2003). In contrast, epigenetic profiling of cancers is less well established. From 2006 to 2013 a number of studies emerged which systematically identified long range epigenetic domains using expression datasets for bladder (Stransky et al., 2006), prostate (Coolen et al., 2010) and breast cancer genomes(Bert et al., 2013; Hsu et al., 2010)(Hsu et al., 2010; Stransky et al., 2006)(Hsu et al., 2010; Stransky et al., 2006). Furthermore domains subject to LRES in cancer have been associated with different disease pathways correlated with aggressiveness of tumours (Dallosso et al., 2012; Vallot et al., 2010).

The Radvanyi lab showed that chromosomal domains containing co-expressed genes could be identified using transcriptome data with positional information to visualise transcriptional correlation of neighbouring genes as maps (Reyal et al., 2005). These

algorithms were utilised in bladder carcinoma to identify LRES domains by categorising the co-regulated gene clusters into those that were copy number dependent and those that were copy number independent (Stransky et al., 2006). Although this method identifies regions of transcription bias that are not attributable to genomic effects it does not automatically mean they are subject to LRES. Indeed this method would identify all possible Regions of Epigenetic Regulation (RERs). It was testing of one such region that was silenced in correlation with changes to the epigenetic landscape that led to the classification of these domains as LRES in bladder carcinoma. Indeed the group later went onto show that infact Multiple Regional Epigenetic Silencing (MRES) occurred at several of these domains where they were epigenetically repressed in association with an aggressive pathway of bladder carcinoma progression (Vallot et al., 2010). Neither study however described regions that were subject to long range epigenetic activation or that were co-ordinately regulated between normal and cancer samples (no change), though they would have been identified by this method.

On the other hand a systematic approach to identify LRES in prostate cancer did look specifically for regions of the genome, within a specified genome size window, that contained genes with low expression levels (Coolen et al., 2010). This approach is exclusive to identifying LRES domains and the study suggested that these changes were attributed once again to changes in the epigenetic landscape where in some cases the domain had hypermethylated DNA, in others a gain of repressive marks and loss of active histone marks, but also domains that were a combination of all three. Interestingly the Clark lab also recently published a study describing long range epigenetically activated regions (LREA) in prostate cancer which were identified by the same method with the exception that they were looking for windows of high gene expression in the transcriptome data (Bert et al., 2012)(Bert et al., 2012)(Bert et al., 2013). As before, investigations into the epigenetic mechanism behind LREA pointed towards chromatin remodelling but also strongly towards DNA hypermethylation at CpG islands or "CpG-island borders" but not at promoters.

The final method I wish to discuss was utilised on estrogen-induced breast cancer cells to provide a map of LRES events in association with this hormone (Hsu et al., 2010). By

integrated analysis of transcriptome, methylome, and ERa binding datasets 11 large domains were identified which were repressed in response to estrogen signalling. Again, as with the studies on bladder (Stransky et al., 2006) and prostate cancer (Bert et al., 2013; Coolen et al., 2010), this analysis focussed on long range epigenetic repression. A number of other studies also used genome-wide methylome data to identify LRES regions in cancer (Dallosso et al., 2012, 2009; Devaney et al., 2011; Frigola et al., 2006; Park et al., 2011).

For my project I was interested in systematically identifying all regions of epigenetic regulation in breast cancer using publicly available datasets. I felt the most comprehensive way to do this was using a method akin to the Radvanyi labs identification of regions of transcriptome bias (Reyal et al., 2005) that are copy number independent (Stransky et al., 2006). This was because I believe this method was the only one that would identify all regions of possible epigenetic regulation – as transcriptome bias could mean that the genes have a co-ordinate up- or down- regulated gene expression pattern. As I was only interested in copy-number independent regions of transcriptome bias, in my methods I first removed all genes that were affected by a copy number aberration in those effected samples. Therefore using the sliding window approach described in Chapter 3 I was then able to assign significance to how well a gene was expressed compared to its neighbours. This method was successfully employed to identify 45 RERs in breast tumours and 71 RERs in cell lines. Using a cell line model I was able to test the mechanism of RER at one such region that overlapped between the tumour and cell line analyses. Later in my PhD this computational approach was also applied to hepatoblastoma as part of a collaboration with the Buendia lab where I identified 68 RERs in this tumour type, which are currently being mechanistically investigated.

7.2: Regions of Epigenetic Regulation

Clusters of co-ordinately regulated genes are not uncommon in the genome and frequently housekeeping genes are organised in this way. Examples of co-ordinately

regulated clusters include the Hox, α/β-globin, MHC, histone and olfactory gene clusters (Sproul et al., 2005). For the simultaneous regulation of multiple genes changes in the chromatin architecture have been implicated. For example the looping of gene clusters to distal control regions containing regulatory elements has been shown for the β-globin locus, HoxD gene cluster and also the MHC clusters (as reviewed in Sproul et al 2005). Changes in the large scale chromatin architecture have also been shown to regulate contiguous genes (e.g. at the HoxB locus) by opening up domains through chromatin decondensation and nuclear reorganisation through looping out of chromosome territories (Chambeyron and Bickmore, 2004). It is thought that chromatin decondensation at such regions and their looping out of their chromosome territories puts them in an environment and state that is permissive to access to transcriptional machinery.

A number of studies have reported that there are large domains (often spanning megabases) that are co-ordinately misregulated in cancers (Bert et al., 2013; Coolen et al., 2010; Dallosso et al., 2012, 2009; Frigola et al., 2006; Hsu et al., 2010; Novak et al., 2006; Stransky et al., 2006). Chromatin structure and organisation within the nucleus would be a good candidate mechanism for regulating multiple neighbouring genes simultaneously. However studies of long-range regulation in cancer have mainly focussed on lower level changes in the epigenetic landscape, namely DNA methylation and histone modifications. One exception to this was for the 14 gene cluster on chromosome 16p11.2 which was reported to be silenced in breast cancer cells in response to estrogen signalling (Hsu et al., 2010). The authors concluded that the region was repressed by gain of the polycomb repressive mark H3K27me3 and DNA hypermethylation at CpG island promoters, but that this repression was reinforced by an "inflexible DNA scaffold" bringing together the 14 gene promoters by DNA looping.

Although this was the first study to show that alterations in chromatin structure itself could contribute to LRES in cancer, I had a number of misgivings about this report. The idea of chromatin forming a permanent loop structure is incompatible with the observations that chromatin is dynamic and continually undergoing changes (Müller et al., 2010). However Hsu et al were describing ER mediated events at this locus in cancer cells and not in normal cells where they suggest the loop structure is more dynamic.

Nevertheless the 16p11.2 region described by that study was also part of a RER domain identified by my own work and, by FISH, I found no significant difference in the level of compaction between ER positive and ER negative cell lines. Furthermore, this subregion (as well as the entire RER domain), showed activated gene expression in tumours compared to normal tissue in my analysis, rather than the repression as described by Hsu et al (2010). By cell line analysis I also showed that this region was relatively activated in ER+ve cell lines compared with ER-ve cell lines using data from Neve et al (2006) (Chapter 4: Figure 4.19). Hsu et al show that this region is repressed in ER+ve cell lines compared to ER-ve cell lines also using the same publically available dataset from Neve et al (2006). Confounded, I manually checked the expression data for genes that are described by Neve et al (2006) as upregulated in luminal/ER+ve cell lines (e.g. *GATA3, SPDEF*) compared to basal/ER-ve cell lines, and also conversely those that are upregulated in basal/ER-ve cell lines compared with luminal/ER+ve cell lines (e.g. *CD44. MSN*), and confirmed that they were behaving as expected in the data I processed from Neve et al (2006). After confirming that my data matched that described by Neve et al (2006) I can only conclude that there has been a technical error in assigning colour to the heat map of gene expression for the 16p11.2 region described as looping in Figure 3B of Hsu et al., (2010). This would however nullify the model of ER mediated looping as a mechanism reinforcing repression of gene expression at this locus. It would however agree with the model of ERα –bound intrachromosomal interactions bringing together genes for transcription in active hubs (Fullwood et al., 2009).

This RER domain at 16p11.2 identified by my analysis was one amongst many regions in breast cancer that served as the first examples of long range epigenetic activation (LREA) during my PhD. LREA was also identified in my analysis of hepatoblastoma tissue data and very recently has also been reported in prostate cancer (Bert et al 2013). Interestingly, many of my RER domains identified were cancer type specific with little overlap between them. Given that I was seeking clusters of genes that were co-ordinately regulated it was not altogether surprising that the RER domains I identified were in gene dense parts of the genome, which often correlated with RIDGEs.

7.3: Chromatin organisation at RER 16p11.2

The RER on chromosome 16p11.2 was used as a candidate for investigating the mechanism of long range control in breast cancer. Apart from the fact that this RER has previously been implicated in LRES in breast cancer cells (Hsu et al., 2010), this was also a good candidate after it appeared in both breast tumour and cell line investigations with significant overlap. Moreover this region was interesting as it was an example of long range activation, which at the time was an unreported phenomena, and it showed a subtype specific gene expression signature.

Chromatin decompaction, as assayed by FISH in this study, has been correlated with transcriptional activation of genes in a number of studies (Chambeyron and Bickmore, 2004; Ragnhild Eskeland et al., 2010; Williamson et al., 2012). By dividing the RER into four subregions it was possible to further refine the RER borders by chromatin compaction analysis combined with computational analysis of transcriptional correlation at smaller window sizes (Chapter 4). Investigation of chromatin compaction at the refined locus (sub-region 2) showed it was relatively decompacted in breast cancer cells and tissues compared to normal breast and that there was also a distinct subtype specific difference with ER+ve cell lines/tumours being the most decompacted overall.

Furthermore comparison of the nuclear localisation of this RER domain showed that it was located significantly more towards the centre of the nucleus in the ER+ve MCF7 cell line compared with the ER-ve MDAMB231 cell line. (This was not however a significant effect for LY2 or MDAMB468). Chromosome 16 on which this RER is located is generally located towards the centre of the nucleus in normal cells (Shelagh Boyle et al., 2001). It is possible that the higher frequency of 16p11.2 localisation with the centre of the nucleus by MCF7 cells compared with MDAMB231 cells was due to the movement of the RER by looping out of its chromosome territory, possibly to a hub of active transcription. The looping out of a chromosome territory coupled with decondensation of chromatin has also been reported for the HoxB locus (Chambeyron and Bickmore, 2004).

7.4: The role of the ER for RER 16p11.2

Having identified a subtype-specific gene expression signature at locus 2 of the RER on 16p11.2 and having shown that it correlated with changes in chromatin organisation I hypothesized that these changes were regulated by the estrogen receptor (ER).

ER+ve tumours, which grow in the presence of estrogen, account for a large proportion of all breast cancers. As discussed in section 1.2.4.6 estrogen action is mediated by the estrogen receptor where ligand binding results in a sequence of events culminating in an upregulation in the expression of estrogen-regulated genes. Blocking estrogen-action using anti-estrogens like Tamoxifen, or estrogen-synthesis with aromatase inhibitors, has been the dominant strategy to treating ER+ve tumours. The exact mechanism by which the estrogen receptor regulates the expression of these genes is unclear but it is thought to involve the recruitment of coregulators, transcription factors and changes to the epigenetic landscape. Moreover ER binding in the breast cancer genome has been shown to alter higher order chromatin architecture allowing ER-bound loci to interact (Carroll et al., 2005; Fullwood et al., 2009; Hsu et al., 2010; Nye et al., 2002).

Investigation of higher order chromatin structure in estrogen-responsive MCF7 cells showed that the level of compaction at these cells was estrogen dependent. In estrogen deprived conditions the locus compacted – an effect that could be reversed by reintroduction of estrogen (Chapter 5). This effect was not apparent in the ER-ve cell line MDAMB231. Given that the level of compaction was found to correlate with the gene expression signature at this locus in breast cancer cell lines and tumours (Chapter 6) these results were a good indication that the estrogen activation of the ER leads to increased chromatin "openness" that allows activation of transcription.

Furthermore, estrogen was found to effect the nuclear localisation of the refined 16p11.2 RER in ER+ve MCF7 cells (Chapter 5). The proportion of FISH signals in the innermost erosion shell was significantly reduced upon estrogen-starvation but increased again after reintroducing estrogen to the media. By contrast in ER-ve MDAMB231 no significant change was seen in the localisation of the RER under any of these conditions. These

results indicated that overall estrogen activation of the ER leads to repositioning of this RER domain possibly into an environment that is enriched in transcription factors. Its possible that the decompaction of the locus coupled with looping out of its chromatin territory (as with HoxB (Chambeyron and Bickmore, 2004; Chambeyron et al., 2005)) in response to estrogen stimulation of the ER is sufficient for aberrant upregulation of these genes in ER+ ve breast cancer. Indeed large-scale decondensation of chromatin in response to ER binding has been previously reported (Nye et al., 2002) though it's precise functional link is currently unknown. During my PhD it certainly appeared that the decondensation of chromatin (as assayed by FISH compaction studies) at locus 2 in concert with its repositioning after estrogen stimulation was enough to allow active transcription, however these two mechanisms have also been shown to function distinctly of one another to activate gene transcription (Morey et al., 2007). Therefore its possible that whilst the decompaction observed in ER+ve cell lines and tumours and indeed upon estrogen stimulation in MCF7 cells facilitates active transcription taking place, that other mechanisms are also required to induce transcription. This could include looping out of a chromosome territory to an active transcription hub and the recruitment of protein complexes associated with active transcription.

In a recent study, estrogen-induced decondensation of chromatin and transcriptional activity was associated with citrullination of histone H3R26 at ER target genes in MCF7 cells (Zhang et al., 2012). Citrullination is catalysed by PAD2 and this enzyme is thought to interact with ERα upon estrogen stimulation creating an open chromatin structure that is permissive of transcription. As discussed in the introduction to this thesis (see section 1.2.3 and 1.3.2) a number of other histone modifications have been associated with altered packing of chromatin, notably hyperacetylation of histones and trimethylation of H3K27me3. Therefore its possible that whilst changes higher order chromatin organisation are key to RER domain formation the repression/activation at such domains is orchestrated by a whole host of transcription factor binding and recruitment of histone modifying enzymes which all work together on a template whose accessibility is facilitated or occluded by the higher order reorganisation of the breast cancer genome.

7.5: The future of RER 16p11.2

In terms of immediate future work it will be important to determine the expression status of 16p11.2 genes upon hormone deprivation and induction by estrogen. This is important to understand why the genes in the RER are misregulated in breast cancer. Whilst in this thesis I have demonstrated that the changes in chromatin compaction and nuclear positioning of this locus are estrogen-mediated, it will be crucial to determine whether the estrogen-mediated decompaction and repositioning of the RER on 16p11.2 in ER+ve cells also leads to their upregulation and relative repression under starvation conditions. To further characterise the topology of the region, taking into account the ERα-binding sites present (Fig 5.1: Fullwood et al 2009), it will be interesting to further tile the region with additional FISH probes. This would tell me exactly how different parts of the region are behaving in response to estrogen deprivation and reintroduction. To ask whether the changes in nuclear area upon hormone deprivation that were observed in ER+ve and ER-ve cells it will be important to FACS analyse the cells for changes in cell cycle under these culture conditions. I also think this study would benefit from testing other RER domains, including those that did not show a ER status specific subtype signature, as it will be important to determine their mechanism of long range misregulation – perhaps also changes to chromatin organisation but mediated by other factors e.g. by chromatin regulatory complexes.

7.6: Concluding remarks

Altered chromatin architecture and nuclear morphology as assessed subjectively by light microscopy is used as a crude measure to diagnose breast tumours from biopsies. This does not however reflect our current molecular understanding of chromatin organisation within the nucleus. During my PhD, I have shown that there are changes to the packing of chromatin and also the relative nuclear position of genes that are misregulated in different types of breast cancer. Furthermore I have shown that such changes in chromatin organisation are mediated by estrogen in ER+ve breast cancer. This is the first step

towards understanding how chromatin re-organisation in the nucleus is linked to breast tumour biology. Further work needs to be done to characterise these changes during disease progression as well as during their response to endocrine therapies – in the hope that such findings will lead to improved methods of diagnosis and a better understanding of how the RER domains are misregulated in breast cancer through changes in chromatin architecture.

Chapter 8

References

Allen, N.E., Beral, V., Casabonne, D., Kan, S.W., Reeves, G.K., Brown, A., Green, J., 2009. Moderate alcohol intake and cancer incidence in women. J. Natl. Cancer Inst. 101, 296–305.

Ashworth, A., 2008. A synthetic lethal therapeutic approach: poly(ADP) ribose polymerase inhibitors for the treatment of cancers deficient in DNA double-strand break repair. J. Clin. Oncol. 26, 3785–3790.

Bachman, K.E., Herman, J.G., Corn, P.G., Merlo, A., Costello, J.F., Cavenee, W.K., Baylin, S.B., Graff, J.R., 1999. Methylation-associated silencing of the tissue inhibitor of metalloproteinase-3 gene suggest a suppressor role in kidney, brain, and other human cancers. Cancer Res. 59, 798–802.

Bal, A., Verma, S., Joshi, K., Singla, A., Thakur, R., Arora, S., Singh, G., 2012. BRCA1-methylated sporadic breast cancers are BRCA-like in showing a basal phenotype and absence of ER expression. Virchows Arch. 461, 305–312.

Banerji, S., Cibulskis, K., Rangel-Escareno, C., Brown, K.K., Carter, S.L., Frederick, A.M., Lawrence, M.S., Sivachenko, A.Y., Sougnez, C., Zou, L., Cortes, M.L., Fernandez-Lopez, J.C., Peng, S., Ardlie, K.G., Auclair, D., Bautista-Piña, V., Duke, F., Francis, J., Jung, J., Maffuz-Aziz, A., Onofrio, R.C., Parkin, M., Pho, N.H., Quintanar-Jurado, V., Ramos, A.H., Rebollar-Vega, R., Rodriguez-Cuevas, S., Romero-Cordoba, S.L., Schumacher, S.E., Stransky, N., Thompson, K.M., Uribe-Figueroa, L., Baselga, J., Beroukhim, R., Polyak, K., Sgroi, D.C., Richardson, A.L., Jimenez-Sanchez, G., Lander, E.S., Gabriel, S.B., Garraway, L.A., Golub, T.R., Melendez-Zajgla, J., Toker, A., Getz, G., Hidalgo-Miranda, A., Meyerson, M., 2012. Sequence analysis of mutations and translocations across breast cancer subtypes. Nature 486, 405–409.

Bantignies, F., Roure, V., Comet, I., Leblanc, B., Schuettengruber, B., Bonnet, J., Tixier, V., Mas, A., Cavalli, G., 2011. Polycomb-dependent regulatory contacts between distant Hox loci in Drosophila. Cell 144, 214–226.

Barber, L.J., Sandhu, S., Chen, L., Campbell, J., Kozarewa, I., Fenwick, K., Assiotis, I., Rodrigues, D.N., Filho, J.S.R., Moreno, V., Mateo, J., Molife, L.R., De Bono, J., Kaye, S., Lord, C.J., Ashworth, A., 2013. Secondary mutations in BRCA2 associated with clinical resistance to a PARP inhibitor. J. Pathol. 229, 422–429.

Barrero, M.J., Malik, S., 2006. Two functional modes of a nuclear receptor-recruited arginine methyltransferase in transcriptional activation. Mol. Cell 24, 233–243.

Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., Zhao, K., 2007. High-resolution profiling of histone methylations in the human genome. Cell 129, 823–837.

Baute, J., Depicker, A., 2008. Base excision repair and its role in maintaining genome stability. Crit. Rev. Biochem. Mol. Biol. 43, 239–276.

Becker, K.F., Atkinson, M.J., Reich, U., Becker, I., Nekarda, H., Siewert, J.R., Höfler, H., 1994. E-cadherin gene mutations provide clues to diffuse type gastric carcinomas. Cancer Res. 54, 3845–3852.

Belandia, B., Parker, M.G., 2003. Nuclear receptors: a rendezvous for chromatin remodeling factors. Cell 114, 277–280.

Belinsky, S.A., Klinge, D.M., Stidley, C.A., Issa, J.-P., Herman, J.G., March, T.H., Baylin, S.B., 2003. Inhibition of DNA methylation and histone deacetylation prevents murine lung cancer. Cancer Res. 63, 7089–7093.

Bell, D.W., Varley, J.M., Szydlo, T.E., Kang, D.H., Wahrer, D.C., Shannon, K.E., Lubratovich, M., Verselis, S.J., Isselbacher, K.J., Fraumeni, J.F., Birch, J.M., Li, F.P., Garber, J.E., Haber, D.A., 1999. Heterozygous germ line hCHK2 mutations in Li-Fraumeni syndrome. Science 286, 2528–2531.

Bell, O., Conrad, T., Kind, J., Wirbelauer, C., Akhtar, A., Schübeler, D., 2008. Transcription-Coupled Methylation of Histone H3 at Lysine 36 Regulates Dosage Compensation by Enhancing Recruitment of the MSL Complex in Drosophila melanogaster. Mol. Cell. Biol. 28, 3401–3409.

Ben-Elazar, S., Yakhini, Z., Yanai, I., 2013. Spatial localization of co-regulated genes exceeds genomic gene clustering in the Saccharomyces cerevisiae genome. Nucleic Acids Res. 41, 2191–2201.

Berger, S.L., Kouzarides, T., Shiekhattar, R., Shilatifard, A., 2009. An operational definition of epigenetics. Genes Dev. 23, 781–783.

Berman, B.P., Weisenberger, D.J., Aman, J.F., Hinoue, T., Ramjan, Z., Liu, Y., Noushmehr, H., Lange, C.P.E., van Dijk, C.M., Tollenaar, R.A.E.M., Van Den Berg, D., Laird, P.W., 2012. Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. Nat. Genet. 44, 40–46.

Bernardino, J., Roux, C., Almeida, A., Vogt, N., Gibaud, A., Gerbault-Seureau, M., Magdelenat, H., Bourgeois, C.A., Malfoy, B., Dutrillaux, B., 1997. DNA hypomethylation in breast cancer: an independent parameter of tumor progression? Cancer Genet. Cytogenet. 97, 83–89.

Bernstein, B.E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D.K., Huebert, D.J., McMahon, S., Karlsson, E.K., Kulbokas, E.J., 3rd, Gingeras, T.R., Schreiber, S.L., Lander, E.S., 2005. Genomic maps and comparative analysis of histone modifications in human and mouse. Cell 120, 169–181.

Bert, S.A., Robinson, M.D., Strbenac, D., Statham, A.L., Song, J.Z., Hulf, T., Sutherland, R.L., Coolen, M.W., Stirzaker, C., Clark, S.J., 2012. Regional Activation of the Cancer Genome by Long-Range Epigenetic Remodeling. Cancer Cell.

Bert, S.A., Robinson, M.D., Strbenac, D., Statham, A.L., Song, J.Z., Hulf, T., Sutherland, R.L., Coolen, M.W., Stirzaker, C., Clark, S.J., 2013. Regional Activation of the Cancer Genome by Long-Range Epigenetic Remodeling. Cancer Cell 23, 9–22.

Bertos, N.R., Park, M., 2011. Breast cancer — one term, many entities? J Clin Invest 121, 3789–3796.

Bhan, A., Hussain, I., Ansari, K.I., Kasiri, S., Bashyal, A., Mandal, S.S., 2013. Antisense Transcript Long Noncoding RNA (lncRNA) HOTAIR is Transcriptionally Induced by Estradiol. J. Mol. Biol.

Bian, Q., Belmont, A.S., 2012. Revisiting higher-order and large-scale chromatin organization. Curr. Opin. Cell Biol. 24, 359–366.

Bird, A., 2007. Perceptions of epigenetics. Nature 447, 396–398.

Bloom, H.J.G., Richardson, W.W., 1957. Histological Grading and Prognosis in Breast Cancer. Br J Cancer 11, 359–377.

Blows, F.M., Driver, K.E., Schmidt, M.K., Broeks, A., van Leeuwen, F.E., Wesseling, J., Cheang, M.C., Gelmon, K., Nielsen, T.O., Blomqvist, C., Heikkilä, P., Heikkinen, T., Nevanlinna, H., Akslen, L.A., Bégin, L.R., Foulkes, W.D., Couch, F.J., Wang, X., Cafourek, V., Olson, J.E., Baglietto, L., Giles, G.G., Severi, G., McLean, C.A., Southey, M.C., Rakha, E., Green, A.R., Ellis, I.O., Sherman, M.E., Lissowska, J., Anderson, W.F., Cox, A., Cross, S.S., Reed, M.W.R., Provenzano, E., Dawson, S.-J., Dunning, A.M., Humphreys, M., Easton, D.F., García-Closas, M., Caldas, C., Pharoah, P.D., Huntsman, D., 2010. Subtyping of Breast Cancer by Immunohistochemistry to Investigate a Relationship between Subtype and Short and Long Term Survival: A Collaborative Analysis of Data for 10,159 Cases from 12 Studies. PLoS Med 7, e1000279.

Botana-Rial, M., de Chiara, L., Valverde, D., Leiro-Fernández, V., Represas-Represas, C., Del Campo-Pérez, V., Fernández-Villar, A., 2012. Prognostic value of aberrant hypermethylation in pleural effusion of lung adenocarcinoma. Cancer Biol. Ther. 13.

Boyd, N.F., Martin, L.J., Bronskill, M., Yaffe, M.J., Duric, N., Minkin, S., 2010. Breast tissue composition and susceptibility to breast cancer. J. Natl. Cancer Inst. 102, 1224–1237.

Boyle, S., Gilchrist, S., Bridger, J.M., Mahy, N.L., Ellis, J.A., Bickmore, W.A., 2001. The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. Hum. Mol. Genet 10, 211–219.

Boyle, S., Gilchrist, S., Bridger, J.M., Mahy, N.L., Ellis, J.A., Bickmore, W.A., 2001. The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. Hum. Mol. Genet. 10, 211–219.

Boyle, S., Rodesch, M.J., Halvensleben, H.A., Jeddeloh, J.A., Bickmore, W.A., 2011. Fluorescence in situ hybridization with high-complexity repeat-free oligonucleotide probes generated by massively parallel synthesis. Chromosome Res. 19, 901–909.

Bracken, A.P., Pasini, D., Capra, M., Prosperini, E., Colli, E., Helin, K., 2003. EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. EMBO J 22, 5323–5335.

Branco, M.R., Pombo, A., 2006. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. PLoS Biol. 4, e138.

Breast cancer and breastfeeding: collaborative reanalysis of individual data from 47 epidemiological studies in 30 countries, including 50302 women with breast cancer and 96973 women without the disease, 2002. . Lancet 360, 187–195.

Breast cancer and hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52,705 women with breast cancer and 108,411 women without breast cancer. Collaborative Group on Hormonal Factors in Breast Cancer, 1997. . Lancet 350, 1047–1059.

Broeks, A., Schmidt, M.K., Sherman, M.E., Couch, F.J., Hopper, J.L., et al., 2011. Low penetrance breast cancer susceptibility loci are associated with specific breast tumor subtypes: findings from the Breast Cancer Association Consortium. Hum. Mol. Genet. 20, 3289–3303.

Brown, J.M., Leach, J., Reittie, J.E., Atzberger, A., Lee-Prudhoe, J., Wood, W.G., Higgs, D.R., Iborra, F.J., Buckle, V.J., 2006. Coregulated human globin genes are frequently in spatial proximity when active. J. Cell Biol. 172, 177–187.

Bundred, N., Gardovskis, J., Jaskiewicz, J., Eglitis, J., Paramonov, V., McCormack, P., Swaisland, H., Cavallin, M., Parry, T., Carmichael, J., Dixon, J.M., 2013. Evaluation of the pharmacodynamics and pharmacokinetics of the PARP inhibitor olaparib: a Phase I multicentre trial in patients scheduled for elective breast cancer surgery. Invest New Drugs.

Cairo, S., Armengol, C., Buendia, M.A., 2012. Activation of Wnt and Myc signaling in hepatoblastoma. Front Biosci (Elite Ed) 4, 480–486.

Cairo, S., Armengol, C., De Reyniès, A., Wei, Y., Thomas, E., Renard, C.-A., Goga, A., Balakrishnan, A., Semeraro, M., Gresh, L., Pontoglio, M., Strick-Marchand, H., Levillayer, F., Nouet, Y., Rickman, D., Gauthier, F., Branchereau, S., Brugières, L.,

Laithier, V., Bouvier, R., Boman, F., Basso, G., Michiels, J.-F., Hofman, P., Arbez-Gindre, F., Jouan, H., Rousselet-Chapeau, M.-C., Berrebi, D., Marcellin, L., Plenat, F., Zachar, D., Joubert, M., Selves, J., Pasquier, D., Bioulac-Sage, P., Grotzer, M., Childs, M., Fabre, M., Buendia, M.-A., 2008. Hepatic stem-like phenotype and interplay of Wnt/beta-catenin and Myc signaling in aggressive childhood liver cancer. Cancer Cell 14, 471–484.

Cameron, E.E., Bachman, K.E., Myöhänen, S., Herman, J.G., Baylin, S.B., 1999. Synergy of demethylation and histone deacetylase inhibition in the re-expression of genes silenced in cancer. Nat. Genet. 21, 103–107.

Canals-Hamann, A.Z., das Neves, R.P., Reittie, J.E., Iñiguez, C., Soneji, S., Enver, T., Buckle, V.J., Iborra, F.J., 2013. A biophysical model for transcription factories. BMC Biophys 6, 2.

Carroll, A.R., Coleman, R.L., Sood, A.K., 2011. Therapeutic advances in women's cancers. Front Biosci (Schol Ed) 3, 82–97.

Carroll, J.S., Liu, X.S., Brodsky, A.S., Li, W., Meyer, C.A., Szary, A.J., Eeckhoute, J., Shao, W., Hestermann, E.V., Geistlinger, T.R., Fox, E.A., Silver, P.A., Brown, M., 2005. Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. Cell 122, 33–43.

Carroll, J.S., Meyer, C.A., Song, J., Li, W., Geistlinger, T.R., Eeckhoute, J., Brodsky, A.S., Keeton, E.K., Fertuck, K.C., Hall, G.F., Wang, Q., Bekiranov, S., Sementchenko, V., Fox, E.A., Silver, P.A., Gingeras, T.R., Liu, X.S., Brown, M., 2006. Genome-wide analysis of estrogen receptor binding sites. Nat. Genet 38, 1289–1297.

Cazzaniga, M., Bonanni, B., 2012. Breast cancer chemoprevention: old and new approaches. J. Biomed. Biotechnol. 2012, 985620.

Ceschin, D.G., Walia, M., Wenk, S.S., Duboé, C., Gaudon, C., Xiao, Y., Fauquier, L., Sankar, M., Vandel, L., Gronemeyer, H., 2011. Methylation specifies distinct estrogen-induced binding site repertoires of CBP to chromatin. Genes Dev. 25, 1132–1146.

Chambeyron, S., Bickmore, W.A., 2004. Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. Genes & Development 18, 1119–1130.

Chambeyron, S., Da Silva, N.R., Lawson, K.A., Bickmore, W.A., 2005. Nuclear re-organisation of the Hoxb complex during mouse embryonic development. Development 132, 2215–2223.

Chan, C.S., Song, J.S., 2008. CCCTC-binding factor confines the distal action of estrogen receptor. Cancer Res. 68, 9041–9049.

Chen, D.-T., Nasir, A., Culhane, A., Venkataramu, C., Fulp, W., Rubio, R., Wang, T., Agrawal, D., McCarthy, S.M., Gruidl, M., Bloom, G., Anderson, T., White, J., Quackenbush, J., Yeatman, T., 2009. Proliferative genes dominate malignancy-risk gene signature in histologically-normal breast tissue. Breast Cancer Res Treat 119, 335–346.

Cheung, P., Allis, C.D., Sassone-Corsi, P., 2000. Signaling to chromatin through histone modifications. Cell 103, 263–271.

Chin, K., DeVries, S., Fridlyand, J., Spellman, P.T., Roydasgupta, R., Kuo, W.-L., Lapuk, A., Neve, R.M., Qian, Z., Ryder, T., Chen, F., Feiler, H., Tokuyasu, T., Kingsley, C., Dairkee, S., Meng, Z., Chew, K., Pinkel, D., Jain, A., Ljung, B.M., Esserman, L., Albertson, D.G., Waldman, F.M., Gray, J.W., 2006. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. Cancer Cell 10, 529–541.

Chisholm, K.M., Wan, Y., Li, R., Montgomery, K.D., Chang, H.Y., West, R.B., 2012. Detection of long non-coding RNA in archival tissue: correlation with polycomb protein expression in primary and metastatic breast carcinoma. PLoS ONE 7, e47998.

Chlebowski, R.T., Blackburn, G.L., Thomson, C.A., Nixon, D.W., Shapiro, A., Hoy, M.K., Goodman, M.T., Giuliano, A.E., Karanja, N., McAndrew, P., Hudis, C., Butler, J., Merkel, D., Kristal, A., Caan, B., Michaelson, R., Vinciguerra, V., Del Prete, S., Winkler, M., Hall, R., Simon, M., Winters, B.L., Elashoff, R.M., 2006. Dietary fat reduction and breast cancer outcome: interim efficacy results from the Women's Intervention Nutrition Study. J. Natl. Cancer Inst. 98, 1767–1776.

Christova, R., Jones, T., Wu, P.-J., Bolzer, A., Costa-Pereira, A.P., Watling, D., Kerr, I.M., Sheer, D., 2007. P-STAT1 mediates higher-order chromatin remodelling of the human MHC in response to IFNgamma. J. Cell. Sci. 120, 3262–3270.

Chu, D., Lu, J., 2008. Novel therapies in breast cancer: what is new from ASCO 2008. Journal of Hematology & Oncology 1, 16.

Chung, C.H., Bernard, P.S., Perou, C.M., 2002. Molecular portraits and the family tree of cancer. Nat Genet 32 Suppl, 533–540.

Cirillo, L.A., Lin, F.R., Cuesta, I., Friedman, D., Jarnik, M., Zaret, K.S., 2002. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. Mol. Cell 9, 279–289.

Cirillo, L.A., McPherson, C.E., Bossard, P., Stevens, K., Cherian, S., Shim, E.Y., Clark, K.L., Burley, S.K., Zaret, K.S., 1998. Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. EMBO J. 17, 244–254.

Clark, G.M., McGuire, W.L., 1991. Follow-up study of HER-2/neu amplification in primary breast cancer. Cancer Res. 51, 944–948.

Clowney, E.J., LeGros, M.A., Mosley, C.P., Clowney, F.G., Markenskoff-Papadimitriou, E.C., Myllys, M., Barnea, G., Larabell, C.A., Lomvardas, S., 2012. Nuclear Aggregation of Olfactory Receptor Genes Governs Their Monogenic Expression. Cell 151, 724–737.

Coolen, M.W., Stirzaker, C., Song, J.Z., Statham, A.L., Kassir, Z., Moreno, C.S., Young, A.N., Varma, V., Speed, T.P., Cowley, M., Lacaze, P., Kaplan, W., Robinson, M.D., Clark, S.J., 2010. Consolidation of the cancer genome into domains of repressive chromatin by long-range epigenetic silencing (LRES) reduces transcriptional plasticity. Nat. Cell Biol 12, 235–246.

Coronado, G.D., Beasley, J., Livaudais, J., 2011. Alcohol consumption and the risk of breast cancer. Salud Publica Mex 53, 440–447.

Cremer, M., Küpper, K., Wagler, B., Wizelman, L., von Hase, J., Weiland, Y., Kreja, L., Diebold, J., Speicher, M.R., Cremer, T., 2003. Inheritance of gene density-related higher order chromatin arrangements in normal and tumor cell nuclei. J. Cell Biol 162, 809–820.

Cremer, T., Cremer, M., Dietzel, S., Müller, S., Solovei, I., Fakan, S., 2006. Chromosome territories--a functional nuclear landscape. Curr. Opin. Cell Biol. 18, 307–316.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., Boyer, L.A., Young, R.A., Jaenisch, R., 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. U.S.A. 107, 21931–21936.

Croft, J.A., Bridger, J.M., Boyle, S., Perry, P., Teague, P., Bickmore, W.A., 1999. Differences in the Localization and Morphology of Chromosomes in the Human Nucleus. J Cell Biol 145, 1119–1131.

Cruickshank, M.N., Besant, P., Ulgiati, D., 2010. The impact of histone post-translational modifications on developmental gene regulation. Amino Acids 39, 1087–1105.

Curtis, C., Shah, S.P., Chin, S.-F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., Gräf, S., Ha, G., Haffari, G., Bashashati, A., Russell, R., McKinney, S., Langerød, A., Green, A., Provenzano, E., Wishart, G., Pinder, S., Watson, P., Markowetz, F., Murphy, L., Ellis, I., Purushotham, A., Børresen-Dale, A.-L., Brenton, J.D., Tavaré, S., Caldas, C., Aparicio, S., 2012. The

genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 486, 346–352.

Dallosso, A.R., Hancock, A.L., Szemes, M., Moorwood, K., Chilukamarri, L., Tsai, H.-H., Sarkar, A., Barasch, J., Vuononvirta, R., Jones, C., Pritchard-Jones, K., Royer-Pokora, B., Lee, S.B., Owen, C., Malik, S., Feng, Y., Frank, M., Ward, A., Brown, K.W., Malik, K., 2009. Frequent long-range epigenetic silencing of protocadherin gene clusters on chromosome 5q31 in Wilms' tumor. PLoS Genet. 5, e1000745.

Dallosso, A.R., Oster, B., Greenhough, A., Thorsen, K., Curry, T.J., Owen, C., Hancock, A.L., Szemes, M., Paraskeva, C., Frank, M., Andersen, C.L., Malik, K., 2012. Long-range epigenetic silencing of chromosome 5q31 protocadherins is involved in early and late stages of colorectal tumorigenesis through modulation of oncogenic pathways. Oncogene.

Dawson, M.A., Foster, S.D., Bannister, A.J., Robson, S.C., Hannah, R., Wang, X., Xhemalce, B., Wood, A.D., Green, A.R., Göttgens, B., Kouzarides, T., 2012. Three Distinct Patterns of Histone H3Y41 Phosphorylation Mark Active Genes. Cell Rep.

De Smet, C., Loriot, A., 2013. DNA hypomethylation and activation of germline-specific genes in cancer. Adv. Exp. Med. Biol. 754, 149–166.

De Wit, E., de Laat, W., 2012. A decade of 3C technologies: insights into nuclear organization. Genes Dev. 26, 11–24.

Deaton, A.M., Bird, A., 2011. CpG islands and the regulation of transcription. Genes Dev. 25, 1010–1022.

Dekker, J., Rippe, K., Dekker, M., Kleckner, N., 2002. Capturing Chromosome Conformation. Science 295, 1306–1311.

Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P.D., Dean, A., Blobel, G.A., 2012. Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. Cell 149, 1233–1244.

Devaney, J., Stirzaker, C., Qu, W., Song, J.Z., Statham, A.L., Patterson, K.I., Horvath, L.G., Tabor, B., Coolen, M.W., Hulf, T., Kench, J.G., Henshall, S.M., Pe Benito, R., Haynes, A.-M., Mayor, R., Peinado, M.A., Sutherland, R.L., Clark, S.J., 2011. Epigenetic deregulation across chromosome 2q14.2 differentiates normal from prostate cancer and provides a regional panel of novel DNA methylation cancer biomarkers. Cancer Epidemiol. Biomarkers Prev. 20, 148–159.

Dey, P., Ponnusamy, M.P., Deb, S., Batra, S.K., 2011. Human RNA Polymerase II-Association Factor 1 (hPaf1/PD2) Regulates Histone Methylation and Chromatin Remodeling in Pancreatic Cancer. PLoS ONE 6, e26926.

Dimri, G., Band, H., Band, V., 2005. Mammary epithelial cell transformation: insights from cell culture and mouse models. Breast Cancer Res. 7, 171–179.

Dixon, J.M., Love, C.D., Renshaw, L., Bellamy, C., Cameron, D.A., Miller, W.R., Leonard, R.C., 1999. Lessons from the use of aromatase inhibitors in the neoadjuvant setting. Endocr. Relat. Cancer 6, 227–230.

Dixon, J.M., Renshaw, L., Bellamy, C., Stuart, M., Hoctin-Boes, G., Miller, W.R., 2000. The effects of neoadjuvant anastrozole (Arimidex) on tumor volume in postmenopausal women with breast cancer: a randomized, double-blind, single-center study. Clin. Cancer Res. 6, 2229–2235.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., Ren, B., 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485, 376–380.

Dobrovic, A., Simpfendorfer, D., 1997. Methylation of the BRCA1 gene in sporadic breast cancer. Cancer Res. 57, 3347–3350.

Dobrzycka, K.M., Townson, S.M., Jiang, S., Oesterreich, S., 2003. Estrogen receptor corepressors -- a role in human breast cancer? Endocr. Relat. Cancer 10, 517–536.

Dong, Q., Wang, D., Bandyopadhyay, A., Gao, H., Gorena, K.M., Hildreth, K., Rebel, V.I., Walter, C.A., Huang, C., Sun, L.-Z., 2013. Mammospheres from murine mammary stem cell-enriched basal cells: Clonal characteristics and repopulating potential. Stem Cell Res 10, 396–404.

Dostie, J., Bickmore, W.A., 2012. Chromosome organization in the nucleus - charting new territory across the Hi-Cs. Curr. Opin. Genet. Dev. 22, 125–131.

Dostie, J., Dekker, J., 2007. Mapping networks of physical interactions between genomic elements using 5C technology. Nat Protoc 2, 988–1002.

Dostie, J., Richmond, T.A., Arnaout, R.A., Selzer, R.R., Lee, W.L., Honan, T.A., Rubio, E.D., Krumm, A., Lamb, J., Nusbaum, C., Green, R.D., Dekker, J., 2006. Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. Genome Res 16, 1299–1309.

Ecker, J.R., Bickmore, W.A., Barroso, I., Pritchard, J.K., Gilad, Y., Segal, E., 2012. Genomics: ENCODE explained. Nature 489, 52–55.

Eden, E., Lipson, D., Yogev, S., Yakhini, Z., 2007. Discovering Motifs in Ranked Lists of DNA Sequences. PLoS Comput Biol 3, e39.

Eden, E., Navon, R., Steinfeld, I., Lipson, D., Yakhini, Z., 2009. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics 10, 48.

Eheman, C.R., Shaw, K.M., Ryerson, A.B., Miller, J.W., Ajani, U.A., White, M.C., 2009. The changing incidence of in situ and invasive ductal and lobular breast carcinomas: United States, 1999-2004. Cancer Epidemiol. Biomarkers Prev. 18, 1763–1769.

Ehrlich, M., Gama-Sosa, M.A., Huang, L.-H., Midgett, R.M., Kuo, K.C., McCune, R.A., Gehrke, C., 1982. Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells. Nucl. Acids Res. 10, 2709–2721.

Elenbaas, B., Spirio, L., Koerner, F., Fleming, M.D., Zimonjic, D.B., Donaher, J.L., Popescu, N.C., Hahn, W.C., Weinberg, R.A., 2001. Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells. Genes Dev. 15, 50–65.

Ellis, M.J., Ding, L., Shen, D., Luo, J., Suman, V.J., et al., 2012. Whole-genome analysis informs breast cancer response to aromatase inhibition. Nature 486, 353–360.

Endoh, M., Endo, T.A., Endoh, T., Isono, K., Sharif, J., Ohara, O., Toyoda, T., Ito, T., Eskeland, R., Bickmore, W.A., Vidal, M., Bernstein, B.E., Koseki, H., 2012. Histone H2A mono-ubiquitination is a crucial step to mediate PRC1-dependent repression of developmental genes to maintain ES cell identity. PLoS Genet. 8, e1002774.

Eskeland, R., Freyer, E., Leeb, M., Wutz, A., Bickmore, W.A., 2010. Histone acetylation and the maintenance of chromatin compaction by Polycomb repressive complexes. Cold Spring Harb. Symp. Quant. Biol. 75, 71–78.

Eskeland, R., Leeb, M., Grimes, G.R., Kress, C., Boyle, S., Sproul, D., Gilbert, N., Fan, Y., Skoultchi, A.I., Wutz, A., 2010. Ring1B Compacts Chromatin Structure and Represses Gene Expression Independent of Histone Ubiquitination. Molecular Cell 38, 452–464.

Esteller, M., Corn, P.G., Urena, J.M., Gabrielson, E., Baylin, S.B., Herman, J.G., 1998. Inactivation of glutathione S-transferase P1 gene by promoter hypermethylation in human neoplasia. Cancer Res. 58, 4515–4518.

Fackenthal, J.D., Olopade, O.I., 2007. Breast cancer risk associated with BRCA1 and BRCA2 in diverse populations. Nat. Rev. Cancer 7, 937–948.

Ferguson, A.T., Evron, E., Umbricht, C.B., Pandita, T.K., Chan, T.A., Hermeking, H., Marks, J.R., Lambers, A.R., Futreal, P.A., Stampfer, M.R., Sukumar, S., 2000. High frequency of hypermethylation at the 14-3-3 sigma locus leads to gene silencing in breast cancer. Proc. Natl. Acad. Sci. U.S.A. 97, 6049–6054.

Ferlay, J., Autier, P., Boniol, M., Heanue, M., Colombet, M., Boyle, P., 2007. Estimates of the cancer incidence and mortality in Europe in 2006. Ann. Oncol. 18, 581–592.

Finlan, L.E., Sproul, D., Thomson, I., Boyle, S., Kerr, E., Perry, P., Ylstra, B., Chubb, J.R., Bickmore, W.A., Reik, W., 2008. Recruitment to the Nuclear Periphery Can Alter Expression of Genes in Human Cells. PLoS Genetics 4, e1000039.

Fleeman, N., Bagust, A., Boland, A., Dickson, R., Dundar, Y., Moonan, M., Oyee, J., Blundell, M., Davis, H., Armstrong, A., Thorp, N., 2011. Lapatinib and trastuzumab in combination with an aromatase inhibitor for the first-line treatment of metastatic hormone receptor-positive breast cancer which over-expresses human epidermal growth factor 2 (HER2): a systematic review and economic analysis. Health Technol Assess 15, 1–93, iii–iv.

Flegel, C., Manteniotis, S., Osthold, S., Hatt, H., Gisselmann, G., 2013. Expression Profile of Ectopic Olfactory Receptors Determined by Deep Sequencing. PLoS One 8.

Ford, D., Easton, D.F., Stratton, M., Narod, S., Goldgar, D., Devilee, P., Bishop, D.T., Weber, B., Lenoir, G., Chang-Claude, J., Sobol, H., Teare, M.D., Struewing, J., Arason, A., Scherneck, S., Peto, J., Rebbeck, T.R., Tonin, P., Neuhausen, S., Barkardottir, R., Eyfjord, J., Lynch, H., Ponder, B.A., Gayther, S.A., Zelada-Hedman, M., 1998. Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. The Breast Cancer Linkage Consortium. Am. J. Hum. Genet. 62, 676–689.

Förster, C., Mäkela, S., Wärri, A., Kietz, S., Becker, D., Hultenby, K., Warner, M., Gustafsson, J.-A., 2002. Involvement of estrogen receptor beta in terminal differentiation of mammary gland epithelium. Proc. Natl. Acad. Sci. U.S.A. 99, 15578–15583.

Francis, N.J., Kingston, R.E., Woodcock, C.L., 2004. Chromatin Compaction by a Polycomb Group Protein Complex. Science 306, 1574–1577.

Fridlyand, J., Snijders, A.M., Ylstra, B., Li, H., Olshen, A., Segraves, R., Dairkee, S., Tokuyasu, T., Ljung, B.M., Jain, A.N., McLennan, J., Ziegler, J., Chin, K., Devries, S., Feiler, H., Gray, J.W., Waldman, F., Pinkel, D., Albertson, D.G., 2006. Breast tumor copy number aberration phenotypes and genomic instability. BMC Cancer 6, 96.

Frierson, H.F., Jr, Wolber, R.A., Berean, K.W., Franquemont, D.W., Gaffey, M.J., Boyd, J.C., Wilbur, D.C., 1995. Interobserver reproducibility of the Nottingham modification of the Bloom and Richardson histologic grading scheme for infiltrating ductal carcinoma. Am. J. Clin. Pathol. 103, 195–198.

Frigola, J., Song, J., Stirzaker, C., Hinshelwood, R.A., Peinado, M.A., Clark, S.J., 2006. Epigenetic remodeling in colorectal cancer results in coordinate gene suppression across an entire chromosome band. Nat. Genet 38, 540–549.

Fullwood, M.J., Liu, M.H., Pan, Y.F., Liu, J., Xu, H., Mohamed, Y.B., Orlov, Y.L., Velkov, S., Ho, A., Mei, P.H., Chew, E.G.Y., Huang, P.Y.H., Welboren, W.-J., Han, Y., Ooi, H.S., Ariyaratne, P.N., Vega, V.B., Luo, Y., Tan, P.Y., Choy, P.Y., Wansa, K.D.S.A., Zhao, B., Lim, K.S., Leow, S.C., Yow, J.S., Joseph, R., Li, H., Desai, K.V., Thomsen, J.S., Lee, Y.K., Karuturi, R.K.M., Herve, T., Bourque, G., Stunnenberg, H.G., Ruan, X., Cacheux-Rataboul, V., Sung, W.-K., Liu, E.T., Wei, C.-L., Cheung, E., Ruan, Y., 2009. An oestrogen-receptor-[agr]-bound human chromatin interactome. Nature 462, 58–64.

Fussner, E., Ching, R.W., Bazett-Jones, D.P., 2011. Living without 30 nm chromatin fibers. Trends in Biochemical Sciences 36, 1–6.

Gage, M., Wattendorf, D., Henry, L.R., 2012. Translational advances regarding hereditary breast cancer syndromes. J Surg Oncol 105, 444–451.

Gal-Yam, E.N., Egger, G., Iniguez, L., Holster, H., Einarsson, S., Zhang, X., Lin, J.C., Liang, G., Jones, P.A., Tanay, A., 2008. Frequent switching of Polycomb repressive marks and DNA hypermethylation in the PC3 prostate cancer cell line. Proc. Natl. Acad. Sci. U.S.A. 105, 12979–12984.

Garcia-Bassets, I., Kwon, Y.-S., Telese, F., Prefontaine, G.G., Hutt, K.R., Cheng, C.S., Ju, B.-G., Ohgi, K.A., Wang, J., Escoubet-Lozach, L., Rose, D.W., Glass, C.K., Fu, X.-D., Rosenfeld, M.G., 2007. Histone methylation-dependent mechanisms impose ligand dependency for gene activation by nuclear receptors. Cell 128, 505–518.

Garcia-Closas, M., Chanock, S., 2008. Genetic susceptibility loci for breast cancer by estrogen receptor status. Clin. Cancer Res. 14, 8000–8009.

Garcia-Closas, M., Hall, P., Nevanlinna, H., Pooley, K., Morrison, J., et al., 2008. Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. PLoS Genet. 4, e1000054.

García-Pedrero, J.M., Kiskinis, E., Parker, M.G., Belandia, B., 2006. The SWI/SNF Chromatin Remodeling Subunit BAF57 Is a Critical Regulator of Estrogen Receptor Function in Breast Cancer Cells. J. Biol. Chem. 281, 22656–22664.

Gehani, S.S., Agrawal-Singh, S., Dietrich, N., Christophersen, N.S., Helin, K., Hansen, K., 2010. Polycomb group protein displacement and gene activation through MSK-dependent H3K27me3S28 phosphorylation. Mol. Cell 39, 886–900.

Ghayad, S.E., Cohen, P.A., 2010. Inhibitors of the PI3K/Akt/mTOR pathway: new hope for breast cancer patients. Recent Pat Anticancer Drug Discov 5, 29–57.

Gil, J., Wu, H., Wang, B.Y., 2002. Image analysis and morphometry in the diagnosis of breast cancer. Microsc. Res. Tech. 59, 109–118.

Gilbert, N., Boyle, S., Fiegler, H., Woodfine, K., Carter, N.P., Bickmore, W.A., 2004. Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. Cell 118, 555–566.

Ginestier, C., Hur, M.H., Charafe-Jauffret, E., Monville, F., Dutcher, J., Brown, M., Jacquemier, J., Viens, P., Kleer, C.G., Liu, S., Schott, A., Hayes, D., Birnbaum, D., Wicha, M.S., Dontu, G., 2007. ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. Cell Stem Cell 1, 555–567.

Giordano, S.H., Hortobagyi, G.N., 2003. Inflammatory breast cancer: Clinical progress and the main problems that must be addressed. Breast Cancer Res 5, 284–288.

Goh, Y., Fullwood, M.J., Poh, H.M., Peh, S.Q., Ong, C.T., Zhang, J., Ruan, X., Ruan, Y., 2012. Chromatin Interaction Analysis with Paired-End Tag Sequencing (ChIA-PET) for Mapping Chromatin Interactions and Understanding Transcription Regulation. Journal of Visualized Experiments.

Goldhirsch, A., Colleoni, M., Gelber, R.D., 2002. Endocrine therapy of breast cancer. Ann. Oncol. 13 Suppl 4, 61–68.

Graff, J.R., Herman, J.G., Lapidus, R.G., Chopra, H., Xu, R., Jarrard, D.F., Isaacs, W.B., Pitha, P.M., Davidson, N.E., Baylin, S.B., 1995. E-cadherin expression is silenced by DNA hypermethylation in human breast and prostate carcinomas. Cancer Res. 55, 5195–5199.

Granit, R.Z., Gabai, Y., Hadar, T., Karamansha, Y., Liberman, L., Waldhorn, I., Gat-Viks, I., Regev, A., Maly, B., Darash-Yahana, M., Peretz, T., Ben-Porath, I., 2012. EZH2 promotes a bi-lineage identity in basal-like breast cancer cells. Oncogene.

Green, K.A., Carroll, J.S., 2007. Oestrogen-receptor-mediated transcription and the influence of co-factors and chromatin state. Nat. Rev. Cancer 7, 713–722.

Greene, F.L., Page, D.L., Fleming, I.D., Fritz, A., Balch, C.M., Haller, D.G., Morrow, M., 2002. AJCC Cancer Staging Manual (6th Edition), 6th Edition. ed. Springer.

Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., de Klein, A., Wessels, L., de Laat, W., van Steensel, B., 2008. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature 453, 948–951.

Guo, W., Keckesova, Z., Donaher, J.L., Shibue, T., Tischler, V., Reinhardt, F., Itzkovitz, S., Noske, A., Zürrer-Härdi, U., Bell, G., Tam, W.L., Mani, S.A., van Oudenaarden,

A., Weinberg, R.A., 2012. Slug and Sox9 cooperatively determine the mammary stem cell state. Cell 148, 1015–1028.

Gupta, R.A., Shah, N., Wang, K.C., Kim, J., Horlings, H.M., Wong, D.J., Tsai, M.-C., Hung, T., Argani, P., Rinn, J.L., Wang, Y., Brzoska, P., Kong, B., Li, R., West, R.B., van de Vijver, M.J., Sukumar, S., Chang, H.Y., 2010. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature 464, 1071–1076.

Ha, G., Roth, A., Lai, D., Bashashati, A., Ding, J., Goya, R., Giuliany, R., Rosner, J., Oloumi, A., Shumansky, K., Chin, S.-F., Turashvili, G., Hirst, M., Caldas, C., Marra, M.A., Aparicio, S., Shah, S.P., 2012. Integrative analysis of genome-wide loss of heterozygosity and monoallelic expression at nucleotide resolution reveals disrupted pathways in triple-negative breast cancer. Genome Res. 22, 1995–2007.

Hahn, M.A., Hahn, T., Lee, D.-H., Esworthy, R.S., Kim, B., Riggs, A.D., Chu, F.-F., Pfeifer, G.P., 2008. Methylation of Polycomb Target Genes in Intestinal Cancer Is Mediated by Inflammation. Cancer Res 68, 10280–10289.

Hakim, O., Sung, M.-H., Voss, T.C., Splinter, E., John, S., Sabo, P.J., Thurman, R.E., Stamatoyannopoulos, J.A., de Laat, W., Hager, G.L., 2011. Diverse gene reprogramming events occur in the same spatial clusters of distal regulatory elements. Genome Research 21, 697–706.

Hamiche, A., Schultz, P., Ramakrishnan, V., Oudet, P., Prunell, A., 1996. Linker histone-dependent DNA structure in linear mononucleosomes. J. Mol. Biol. 257, 30–42.

Handoko, L., Xu, H., Li, G., Ngan, C.Y., Chew, E., Schnapp, M., Lee, C.W.H., Ye, C., Ping, J.L.H., Mulawadi, F., Wong, E., Sheng, J., Zhang, Y., Poh, T., Chan, C.S., Kunarso, G., Shahab, A., Bourque, G., Cacheux-Rataboul, V., Sung, W.-K., Ruan, Y., Wei, C.-L., 2011. CTCF-mediated functional chromatin interactome in pluripotent cells. Nat. Genet. 43, 630–638.

Hansen, J.C., Ausio, J., Stanik, V.H., van Holde, K.E., 1989. Homogeneous reconstituted oligonucleosomes, evidence for salt-dependent folding in the absence of histone H1. Biochemistry 28, 9129–9136.

Hansen, K.D., Timp, W., Bravo, H.C., Sabunciyan, S., Langmead, B., McDonald, O.G., Wen, B., Wu, H., Liu, Y., Diep, D., Briem, E., Zhang, K., Irizarry, R.A., Feinberg, A.P., 2011. Increased methylation variation in epigenetic domains across cancer types. Nature Genetics 43, 768–775.

Harewood, L., Schutz, F., Boyle, S., Perry, P., Delorenzi, M., Bickmore, W.A., Reymond, A., 2010. The effect of translocation-induced nuclear reorganization on gene expression. Genome Research 20, 554–564.

Harte, M.T., O'Brien, G.J., Ryan, N.M., Gorski, J.J., Savage, K.I., Crawford, N.T., Mullan, P.B., Harkin, D.P., 2010. BRD7, a Subunit of SWI/SNF Complexes, Binds Directly to BRCA1 and Regulates BRCA1-Dependent Transcription. Cancer Res 70, 2538–2547.

Hawkins, R.D., Hon, G.C., Lee, L.K., Ngo, Q., Lister, R., Pelizzola, M., Edsall, L.E., Kuan, S., Luu, Y., Klugman, S., Antosiewicz-Bourget, J., Ye, Z., Espinoza, C., Agarwahl, S., Shen, L., Ruotti, V., Wang, W., Stewart, R., Thomson, J.A., Ecker, J.R., Ren, B., 2010. Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. Cell Stem Cell 6, 479–491.

Haybittle, J.L., Blamey, R.W., Elston, C.W., Johnson, J., Doyle, P.J., Campbell, F.C., Nicholson, R.I., Griffiths, K., 1982. A prognostic index in primary breast cancer. Br. J. Cancer 45, 361–366.

Heard, E., Bickmore, W., 2007. The ins and outs of gene regulation and chromosome territory organisation. Curr. Opin. Cell Biol 19, 311–316.

Hearle, N., Schumacher, V., Menko, F.H., Olschwang, S., Boardman, L.A., Gille, J.J.P., Keller, J.J., Westerman, A.M., Scott, R.J., Lim, W., Trimbath, J.D., Giardiello, F.M., Gruber, S.B., Offerhaus, G.J.A., de Rooij, F.W.M., Wilson, J.H.P., Hansmann, A., Möslein, G., Royer-Pokora, B., Vogel, T., Phillips, R.K.S., Spigelman, A.D., Houlston, R.S., 2006. Frequency and spectrum of cancers in the Peutz-Jeghers syndrome. Clin. Cancer Res. 12, 3209–3215.

Hebbes, T.R., Thorne, A.W., Crane-Robinson, C., 1988. A direct link between core histone acetylation and transcriptionally active chromatin. EMBO J. 7, 1395–1402.

Heldring, N., Pike, A., Andersson, S., Matthews, J., Cheng, G., Hartman, J., Tujague, M., Ström, A., Treuter, E., Warner, M., Gustafsson, J.-Å., 2007. Estrogen Receptors: How Do They Signal and What Are Their Targets. Physiol Rev 87, 905–931.

Helguero, L.A., Faulds, M.H., Gustafsson, J.-A., Haldosén, L.-A., 2005. Estrogen receptors alfa (ERalpha) and beta (ERbeta) differentially regulate proliferation and apoptosis of the normal murine mammary epithelial cell line HC11. Oncogene 24, 6605–6616.

Herman, J.G., Merlo, A., Mao, L., Lapidus, R.G., Issa, J.P., Davidson, N.E., Sidransky, D., Baylin, S.B., 1995. Inactivation of the CDKN2/p16/MTS1 gene is frequently associated with aberrant DNA methylation in all common human cancers. Cancer Res. 55, 4525–4530.

Herschkowitz, J.I., Simin, K., Weigman, V.J., Mikaelian, I., Usary, J., Hu, Z., Rasmussen, K.E., Jones, L.P., Assefnia, S., Chandrasekharan, S., Backlund, M.G., Yin, Y., Khramtsov, A.I., Bastein, R., Quackenbush, J., Glazer, R.I., Brown, P.H.,

Green, J.E., Kopelovich, L., Furth, P.A., Palazzo, J.P., Olopade, O.I., Bernard, P.S., Churchill, G.A., Van Dyke, T., Perou, C.M., 2007. Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. Genome Biol. 8, R76.

Hewitt, S.C., Harrell, J.C., Korach, K.S., 2005. Lessons in estrogen biology from knockout and transgenic animals. Annu. Rev. Physiol. 67, 285–308.

Hindle, W.H., n.d. A Clinical Guidebook for Women's Primary Health Care Providers, 1999 edition. ed. Springer, New York.

Holliday, R., 1990. Mechanisms for the control of gene activity during development. Biol Rev Camb Philos Soc 65, 431–471.

Holstege, H., Joosse, S.A., van Oostrom, C.T.M., Nederlof, P.M., de Vries, A., Jonkers, J., 2009. High incidence of protein-truncating TP53 mutations in BRCA1-related breast cancer. Cancer Res. 69, 3625–3633.

Hon, G., Wang, W., Ren, B., 2009. Discovery and annotation of functional chromatin signatures in the human genome. PLoS Comput. Biol. 5, e1000566.

Hon, G.C., Hawkins, R.D., Caballero, O.L., Lo, C., Lister, R., Pelizzola, M., Valsesia, A., Ye, Z., Kuan, S., Edsall, L.E., Camargo, A.A., Stevenson, B.J., Ecker, J.R., Bafna, V., Strausberg, R.L., Simpson, A.J., Ren, B., 2012. Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. Genome Res. 22, 246–258.

Hong, L., Schroth, G.P., Matthews, H.R., Yau, P., Bradbury, E.M., 1993. Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 "tail" to DNA. J. Biol. Chem. 268, 305–314.

Horike, S., Cai, S., Miyano, M., Cheng, J.-F., Kohwi-Shigematsu, T., 2005. Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. Nat. Genet. 37, 31–40.

Hsu, P.-Y., Hsu, H.-K., Singer, G.A.C., Yan, P.S., Rodriguez, B.A.T., Liu, J.C., Weng, Y.-I., Deatherage, D.E., Chen, Z., Pereira, J.S., Lopez, R., Russo, J., Wang, Q., Lamartiniere, C.A., Nephew, K.P., Huang, T.H.-M., 2010. Estrogen-mediated epigenetic repression of large chromosomal regions through DNA looping. Genome Res.

Huang, Y., Nayak, S., Jankowitz, R., Davidson, N.E., Oesterreich, S., 2011. Epigenetics in breast cancer: what's new? Breast Cancer Res 13, 225.

Huang, Y., Vasilatos, S.N., Boric, L., Shaw, P.G., Davidson, N.E., 2012. Inhibitors of histone demethylation and histone deacetylation cooperate in regulating gene expression and inhibiting growth in human breast cancer cells. Breast Cancer Res. Treat. 131, 777–789.

Hurtado, A., Holmes, K.A., Ross-Innes, C.S., Schmidt, D., Carroll, J.S., 2011. FOXA1 is a key determinant of estrogen receptor function and endocrine response. Nat. Genet. 43, 27–33.

Hwang, C., Giri, V.N., Wilkinson, J.C., Wright, C.W., Wilkinson, A.S., Cooney, K.A., Duckett, C.S., 2008. EZH2 regulates the transcription of estrogen-responsive genes through association with REA, an estrogen receptor corepressor. Breast Cancer Res. Treat 107, 235–242.

Iborra, F.J., Pombo, A., Jackson, D.A., Cook, P.R., 1996. Active RNA polymerases are localized within discrete transcription "factories' in human nuclei. J Cell Sci 109, 1427–1436.

Illingworth, R.S., Gruenewald-Schneider, U., Webb, S., Kerr, A.R.W., James, K.D., Turner, D.J., Smith, C., Harrison, D.J., Andrews, R., Bird, A.P., 2010. Orphan CpG islands identify numerous conserved promoters in the mammalian genome. PLoS Genet. 6.

Iyer, L.M., Tahiliani, M., Rao, A., Aravind, L., 2009. Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. Cell Cycle 8, 1698–1710.

Jackson, D.A., Hassan, A.B., Errington, R.J., Cook, P.R., 1993. Visualization of focal sites of transcription within human nuclei. EMBO J. 12, 1059–1065.

Järvinen, T.A., Pelto-Huikko, M., Holli, K., Isola, J., 2000. Estrogen receptor beta is coexpressed with ERalpha and PR and associated with nodal status, grade, and proliferation rate in breast cancer. Am. J. Pathol. 156, 29–35.

Javierre, B.M., Rodriguez-Ubreva, J., Al-Shahrour, F., Corominas, M., Graña, O., Ciudad, L., Agirre, X., Pisano, D.G., Valencia, A., Roman-Gomez, J., Calasanz, M.J., Prosper, F., Esteller, M., Gonzalez-Sarmiento, R., Ballestar, E., 2011. Long-range epigenetic silencing associates with deregulation of Ikaros targets in colorectal cancer cells. Mol. Cancer Res. 9, 1139–1151.

Jin, S.-G., Kadam, S., Pfeifer, G.P., 2010. Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. Nucl. Acids Res. 38, e125–e125.

Johansen, K.M., Johansen, J., 2006. Regulation of chromatin structure by histone H3S10 phosphorylation. Chromosome Res. 14, 393–404.

Jones, P.A., Baylin, S.B., 2007. The epigenomics of cancer. Cell 128, 683–692.

Jönsson, G., Staaf, J., Vallon-Christersson, J., Ringnér, M., Holm, K., Hegardt, C., Gunnarsson, H., Fagerholm, R., Strand, C., Agnarsson, B.A., Kilpivaara, O., Luts, L., Heikkilä, P., Aittomäki, K., Blomqvist, C., Loman, N., Malmström, P., Olsson, H., Johannsson, O.T., Arason, A., Nevanlinna, H., Barkardottir, R.B., Borg, A., 2010. Genomic subtypes of breast cancer identified by array-comparative genomic hybridization display distinct molecular and clinical characteristics. Breast Cancer Res 12, R42.

Jozwik, K.M., Carroll, J.S., 2012. Pioneer factors in hormone-dependent cancers. Nat. Rev. Cancer 12, 381–385.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., Taatjes, D.J., Dekker, J., Young, R.A., 2010. Mediator and cohesin connect gene expression and chromatin architecture. Nature 467, 430–435.

Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F., Chen, L., 2012. Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. Nat. Biotechnol. 30, 90–98.

Karp, G., Maissel, A., Livneh, E., 2007. Hormonal regulation of PKC: Estrogen up-regulates PKCη expression in estrogen-responsive breast cancer cells. Cancer Letters 246, 173–181.

Kassabov, S.R., Zhang, B., Persinger, J., Bartholomew, B., 2003. SWI/SNF Unwraps, Slides, and Rewraps the Nucleosome. Molecular Cell 11, 391–403.

Keen, J.C., Yan, L., Mack, K.M., Pettit, C., Smith, D., Sharma, D., Davidson, N.E., 2003. A novel histone deacetylase inhibitor, scriptaid, enhances expression of functional estrogen receptor alpha (ER) in ER negative human breast cancer cells in combination with 5-aza 2'-deoxycytidine. Breast Cancer Res. Treat. 81, 177–186.

Khan, O., La Thangue, N.B., 2008. Drug Insight: histone deacetylase inhibitor-based therapies for cutaneous T-cell lymphomas. Nat Clin Pract Oncol 5, 714–726.

Kinney, S.R.M., Pradhan, S., 2013. Ten eleven translocation enzymes and 5-hydroxymethylation in mammalian development and cancer. Adv. Exp. Med. Biol. 754, 57–79.

Kirova, Y.M., 2010. Recent advances in breast cancer radiotherapy: Evolution or revolution, or how to decrease cardiac toxicity? World J Radiol 2, 103–108.

Kiyuna, T., Saito, A., Kerr, E., Bickmore, W., 2008. Characterization of chromatin texture by contour complexity for cancer cell classification, in: 8th IEEE International

Conference on BioInformatics and BioEngineering, 2008. BIBE 2008. Presented at the 8th IEEE International Conference on BioInformatics and BioEngineering, 2008. BIBE 2008, pp. 1 –6.

Klaassen, U., Seeber, S., 1997. Metastatic breast cancer: treatment with fluorouracil-based combinations. Oncology (Williston Park, N.Y.) 11, 69–73.

Kleer, C.G., Cao, Q., Varambally, S., Shen, R., Ota, I., Tomlins, S.A., Ghosh, D., Sewalt, R.G.A.B., Otte, A.P., Hayes, D.F., Sabel, M.S., Livant, D., Weiss, S.J., Rubin, M.A., Chinnaiyan, A.M., 2003. EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. Proc. Natl. Acad. Sci. U.S.A 100, 11606–11611.

Kleihues, P., Schäuble, B., zur Hausen, A., Estève, J., Ohgaki, H., 1997. Tumors associated with p53 germline mutations: a synopsis of 91 families. Am. J. Pathol. 150, 1–13.

Koboldt, D.C., Fulton, R.S., McLellan, M.D., Schmidt, H., Kalicki-Veizer, J., et al., 2012. Comprehensive molecular portraits of human breast tumours. Nature.

Kriaucionis, S., Heintz, N., 2009. The nuclear DNA base, 5-hydroxymethylcytosine is present in brain and enriched in Purkinje neurons. Science 324, 929–930.

Kristensen, L.S., Nielsen, H.M., Hansen, L.L., 2009. Epigenetics and cancer treatment. European Journal of Pharmacology 625, 131–142.

Lapidus, R.G., Ferguson, A.T., Ottaviano, Y.L., Parl, F.F., Smith, H.S., Weitzman, S.A., Baylin, S.B., Issa, J.P., Davidson, N.E., 1996. Methylation of estrogen and progesterone receptor gene 5' CpG islands correlates with lack of estrogen and progesterone receptor gene expression in breast tumors. Clin. Cancer Res. 2, 805–810.

Larsen, F., Gundersen, G., Lopez, R., Prydz, H., 1992. CpG islands as gene markers in the human genome. Genomics 13, 1095–1107.

Le Romancer, M., Treilleux, I., Leconte, N., Robin-Lespinasse, Y., Sentis, S., Bouchekioua-Bouzaghou, K., Goddard, S., Gobert-Gosse, S., Corbo, L., 2008. Regulation of estrogen rapid signaling through arginine methylation by PRMT1. Mol. Cell 31, 212–221.

Leitman, D.C., Paruthiyil, S., Vivar, O.I., Saunier, E.F., Herber, C.B., Cohen, I., Tagliaferri, M., Speed, T.P., 2010. Regulation of specific target genes and biological responses by estrogen receptor subtype agonists. Curr Opin Pharmacol 10, 629–636.

Li, W., Liu, M., 2011. Distribution of 5-Hydroxymethylcytosine in Different Human Tissues. Journal of Nucleic Acids 2011, 1–5.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., Sandstrom, R., Bernstein, B., Bender, M.A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L.A., Lander, E.S., Dekker, J., 2009. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. Science 326, 289–293.

Lim, S., Janzer, A., Becker, A., Zimmer, A., Schüle, R., Buettner, R., Kirfel, J., 2010. Lysine-specific demethylase 1 (LSD1) is highly expressed in ER-negative breast cancers and a biomarker predicting aggressive biology. Carcinogenesis 31, 512–520.

Lu, S., Singh, K., Mangray, S., Tavares, R., Noble, L., Resnick, M.B., Yakirevich, E., 2012. Claudin expression in high-grade invasive ductal carcinoma of the breast: correlation with the molecular subtype. Mod. Pathol.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., Richmond, T.J., 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 389, 251–260.

Lupien, M., Eeckhoute, J., Meyer, C.A., Wang, Q., Zhang, Y., Li, W., Carroll, J.S., Liu, X.S., Brown, M., 2008. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. Cell 132, 958–970.

Ma, H., Baumann, C.T., Li, H., Strahl, B.D., Rice, R., Jelinek, M.A., Aswad, D.W., Allis, C.D., Hager, G.L., Stallcup, M.R., 2001. Hormone-dependent, CARM1-directed, arginine-specific methylation of histone H3 on a steroid-regulated promoter. Curr. Biol. 11, 1981–1985.

Ma, L., Young, J., Prabhala, H., Pan, E., Mestdagh, P., Muth, D., Teruya-Feldstein, J., Reinhardt, F., Onder, T.T., Valastyan, S., Westermann, F., Speleman, F., Vandesompele, J., Weinberg, R.A., 2010. miR-9, a MYC/MYCN-activated microRNA, regulates E-cadherin and cancer metastasis. Nature Cell Biology.

Marsh, D.J., Coulon, V., Lunetta, K.L., Rocca-Serra, P., Dahia, P.L., Zheng, Z., Liaw, D., Caron, S., Duboué, B., Lin, A.Y., Richardson, A.L., Bonnetblanc, J.M., Bressieux, J.M., Cabarrot-Moreau, A., Chompret, A., Demange, L., Eeles, R.A., Yahanda, A.M., Fearon, E.R., Fricker, J.P., Gorlin, R.J., Hodgson, S.V., Huson, S., Lacombe, D., Eng, C., 1998. Mutation spectrum and genotype-phenotype analyses in Cowden disease and Bannayan-Zonana syndrome, two hamartoma syndromes with germline PTEN mutation. Hum. Mol. Genet. 7, 507–515.

Marson, C.M., 2009. Histone deacetylase inhibitors: design, structure-activity relationships and therapeutic implications for cancer. Anticancer Agents Med Chem 9, 661–692.

Marzluff, W.F., Wagner, E.J., Duronio, R.J., 2008. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. Nat. Rev. Genet. 9, 843–854.

Matsuoka, S., Huang, M., Elledge, S.J., 1998. Linkage of ATM to cell cycle regulation by the Chk2 protein kinase. Science 282, 1893–1897.

Mavaddat, N., Barrowdale, D., Andrulis, I.L., Domchek, S.M., Eccles, D., et al., 2012. Pathology of breast and ovarian cancers among BRCA1 and BRCA2 mutation carriers: results from the Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA). Cancer Epidemiol. Biomarkers Prev. 21, 134–147.

McCormack, V.A., dos Santos Silva, I., 2006. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. Cancer Epidemiol. Biomarkers Prev. 15, 1159–1169.

Meaburn, K.J., Gudla, P.R., Khan, S., Lockett, S.J., Misteli, T., 2009. Disease-specific gene repositioning in breast cancer. J. Cell Biol. 187, 801–812.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., Gnirke, A., Jaenisch, R., Lander, E.S., 2008. Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature 454, 766–770.

Mendes-Pereira, A.M., Sims, D., Dexter, T., Fenwick, K., Assiotis, I., Kozarewa, I., Mitsopoulos, C., Hakas, J., Zvelebil, M., Lord, C.J., Ashworth, A., 2012. Genome-wide functional screen identifies a compendium of genes affecting sensitivity to tamoxifen. Proc. Natl. Acad. Sci. U.S.A. 109, 2730–2735.

Milne, T.A., Briggs, S.D., Brock, H.W., Martin, M.E., Gibbs, D., Allis, C.D., Hess, J.L., 2002. MLL targets SET domain methyltransferase activity to Hox gene promoters. Mol. Cell 10, 1107–1117.

Moen, P.T., Jr, Johnson, C.V., Byron, M., Shopland, L.S., de la Serna, I.L., Imbalzano, A.N., Lawrence, J.B., 2004. Repositioning of muscle-specific genes relative to the periphery of SC-35 domains during skeletal myogenesis. Mol. Biol. Cell 15, 197–206.

Mohsin, S.K., Weiss, H., Havighurst, T., Clark, G.M., Berardo, M., Roanh, L.D., To, T.V., Qian, Zhang, Qian, Zho, Love, R.R., Allred, D.C., 2004. Progesterone receptor by immunohistochemistry and clinical outcome in breast cancer: a validation study. Mod. Pathol. 17, 1545–1554.

Morey, C., Da Silva, N.R., Perry, P., Bickmore, W.A., 2007. Nuclear reorganisation and chromatin decondensation are conserved, but distinct, mechanisms linked to Hox gene activation. Development 134, 909–919.

Morey, C., Kress, C., Bickmore, W.A., 2009. Lack of bystander activation shows that localization exterior to chromosome territories is not sufficient to up-regulate gene expression. Genome Res. 19, 1184–1194.

Morrow, M., Harris, J.R., Schnitt, S.J., 2012. Surgical margins in lumpectomy for breast cancer--bigger is not better. N. Engl. J. Med. 367, 79–82.

Mostoslavsky, R., 2008. DNA repair, insulin signaling and sirtuins: at the crossroads between cancer and aging. Frontiers in Bioscience Volume, 6966.

Müller, I., Boyle, S., Singer, R.H., Bickmore, W.A., Chubb, J.R., 2010. Stable morphology, but dynamic internal reorganisation, of interphase human chromosomes in living cells. PLoS ONE 5, e11560.

Muranen, T.A., Greco, D., Fagerholm, R., Kilpivaara, O., Kämpjärvi, K., Aittomäki, K., Blomqvist, C., Heikkilä, P., Borg, Å., Nevanlinna, H., 2011. Breast tumors from CHEK2 1100delC-mutation carriers: genomic landscape and clinical implications. Breast Cancer Research 13, R90.

Naeem, H., Cheng, D., Zhao, Q., Underhill, C., Tini, M., Bedford, M.T., Torchia, J., 2007. The activity and stability of the transcriptional coactivator p/CIP/SRC-3 are regulated by CARM1-dependent methylation. Mol. Cell. Biol. 27, 120–134.

Narod, S.A., Foulkes, W.D., 2004. BRCA1 and BRCA2: 1994 and beyond. Nature Reviews Cancer 4, 665–676.

Nestor, C., Ruzov, A., Meehan, R., Dunican, D., 2010. Enzymatic approaches and bisulfite sequencing cannot distinguish between 5-methylcytosine and 5-hydroxymethylcytosine in DNA. BioTechniques 48, 317–319.

Nestor, C.E., Ottaviano, R., Reddington, J., Sproul, D., Reinhardt, D., Dunican, D., Katz, E., Dixon, J.M., Harrison, D.J., Meehan, R.R., 2012. Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes. Genome Res. 22, 467–477.

Neuhaus, E.M., Zhang, W., Gelis, L., Deng, Y., Noldus, J., Hatt, H., 2009. Activation of an olfactory receptor inhibits proliferation of prostate cancer cells. J. Biol. Chem.

Neve, R.M., Chin, K., Fridlyand, J., Yeh, J., Baehner, F.L., Fevr, T., Clark, L., Bayani, N., Coppe, J.-P., Tong, F., Speed, T., Spellman, P.T., DeVries, S., Lapuk, A., Wang, N.J., Kuo, W.-L., Stilwell, J.L., Pinkel, D., Albertson, D.G., Waldman, F.M., McCormick, F., Dickson, R.B., Johnson, M.D., Lippman, M., Ethier, S., Gazdar, A., Gray, J.W., 2006. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell 10, 515–527.

Neve, R.M., Chin, K., Fridlyand, J., Yeh, J., Baehner, F.L., Fevr, T., Clark, L., Bayani, N., Coppe, J.-P., Tong, F., Speed, T., Spellman, P.T., DeVries, S., Lapuk, A., Wang, N.J., Kuo, W.-L., Stilwell, J.L., Pinkel, D., Albertson, D.G., Waldman, F.M., McCormick, F., Dickson, R.B., Johnson, M.D., Lippman, M., Ethier, S., Gazdar, A., Gray, J.W., 2006. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell 10, 515–527.

Nilsson, S., Gustafsson, J.-A., 2002. Biological role of estrogen and estrogen receptors. Crit. Rev. Biochem. Mol. Biol. 37, 1–28.

Noma K, Allis, C.D., Grewal, S.I., 2001. Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. Science 293, 1150–1155.

Noordermeer, D., Leleu, M., Splinter, E., Rougemont, J., Laat, W.D., Duboule, D., 2011. The Dynamic Architecture of Hox Gene Clusters. Science 334, 222–225.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., Gribnau, J., Barillot, E., Blüthgen, N., Dekker, J., Heard, E., 2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature 485, 381–385.

Novak, P., Jensen, T., Oshiro, M.M., Watts, G.S., Kim, C.J., Futscher, B.W., 2008. Agglomerative epigenetic aberrations are a common event in human breast cancer. Cancer Res. 68, 8616–8625.

Novak, P., Jensen, T., Oshiro, M.M., Wozniak, R.J., Nouzova, M., Watts, G.S., Klimecki, W.T., Kim, C., Futscher, B.W., 2006. Epigenetic Inactivation of the HOXA Gene Cluster in Breast Cancer. Cancer Research 66, 10664–10670.

Nowak, S.J., Corces, V.G., 2000. Phosphorylation of histone H3 correlates with transcriptionally active loci. Genes Dev. 14, 3003–3013.

Nye, A.C., Rajendran, R.R., Stenoien, D.L., Mancini, M.A., Katzenellenbogen, B.S., Belmont, A.S., 2002. Alteration of large-scale chromatin structure by estrogen receptor. Mol. Cell. Biol 22, 3437–3449.

Ohm, J.E., McGarvey, K.M., Yu, X., Cheng, L., Schuebel, K.E., Cope, L., Mohammad, H.P., Chen, W., Daniel, V.C., Yu, W., Berman, D.M., Jenuwein, T., Pruitt, K., Sharkis, S.J., Watkins, D.N., Herman, J.G., Baylin, S.B., 2007. A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. Nat. Genet 39, 237–242.

Olayioye, M., 2001. Intracellular signaling pathways of ErbB2/HER-2 and family members. Breast Cancer Research 3, 385.

Olins, A.L., Olins, D.E., 1974. Spheroid chromatin units (v bodies). Science 183, 330–332.

Omoto, Y., Inoue, S., Ogawa, S., Toyama, T., Yamashita, H., Muramatsu, M., Kobayashi, S., Iwase, H., 2001. Clinical value of the wild-type estrogen receptor beta expression in breast cancer. Cancer Lett. 163, 207–212.

Osborne, C.S., Chakalova, L., Brown, K.E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J.A., Lopes, S., Reik, W., Fraser, P., 2004. Active genes dynamically colocalize to shared sites of ongoing transcription. Nat Genet 36, 1065–1071.

Osborne, C.S., Chakalova, L., Mitchell, J.A., Horton, A., Wood, A.L., Bolland, D.J., Corcoran, A.E., Fraser, P., 2007. Myc dynamically and preferentially relocates to a transcription factory occupied by Igh. PLoS Biol. 5, e192.

Ottaviano, Y.L., Issa, J.P., Parl, F.F., Smith, H.S., Baylin, S.B., Davidson, N.E., 1994. Methylation of the estrogen receptor gene CpG island marks loss of estrogen receptor expression in human breast cancer cells. Cancer Res. 54, 2552–2555.

Otte, A.P., Kwaks, T.H., 2003. Gene repression by Polycomb group protein complexes: a distinct complex for every occasion? Current Opinion in Genetics & Development 13, 448–454.

Paik, S., Hazan, R., Fisher, E.R., Sass, R.E., Fisher, B., Redmond, C., Schlessinger, J., Lippman, M.E., King, C.R., 1990. Pathologic findings from the National Surgical Adjuvant Breast and Bowel Project: prognostic significance of erbB-2 protein overexpression in primary breast cancer. J. Clin. Oncol. 8, 103–112.

Pal, B., Bouras, T., Shi, W., Vaillant, F., Sheridan, J.M., Fu, N., Breslin, K., Jiang, K., Ritchie, M.E., Young, M., Lindeman, G.J., Smyth, G.K., Visvader, J.E., 2013. Global changes in the mammary epigenome are induced by hormonal cues and coordinated by Ezh2. Cell Rep 3, 411–426.

Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H.C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T., Cobb, B.S., Yokomori, K., Dillon, N., Aragon, L., Fisher, A.G., Merkenschlager, M., 2008. Cohesins functionally associate with CTCF on mammalian chromosome arms. Cell 132, 422–433.

Park, J.-H., Park, J., Choi, J.K., Lyu, J., Bae, M.-G., Lee, Y.-G., Bae, J.-B., Park, D.Y., Yang, H.-K., Kim, T.-Y., Kim, Y.-J., 2011. Identification of DNA methylation changes associated with human gastric cancer. BMC Med Genomics 4, 82.

Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S.W.M., Solovei, I., Brugman, W., Gräf, S., Flicek, P., Kerkhoven, R.M., van Lohuizen, M., Reinders, M., Wessels, L., van Steensel, B., 2010. Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. Mol. Cell 38, 603–613.

Pharoah, P.D., Day, N.E., Caldas, C., 1999. Somatic mutations in the p53 gene and prognosis in breast cancer: a meta-analysis. Br. J. Cancer 80, 1968–1973.

Phillips, J.E., Corces, V.G., 2009. CTCF: Master Weaver of the Genome. Cell 137, 1194–1211.

Pokholok, D.K., Harbison, C.T., Levine, S., Cole, M., Hannett, N.M., Lee, T.I., Bell, G.W., Walker, K., Rolfe, P.A., Herbolsheimer, E., Zeitlinger, J., Lewitter, F., Gifford, D.K., Young, R.A., 2005. Genome-wide map of nucleosome acetylation and methylation in yeast. Cell 122, 517–527.

Powell, E., Shanle, E., Brinkman, A., Li, J., Keles, S., Wisinski, K.B., Huang, W., Xu, W., 2012. Identification of Estrogen Receptor Dimer Selective Ligands Reveals Growth-Inhibitory Effects on Cells That Co-Express ERα and ERβ. PLoS ONE 7, e30993.

Prall, O.W., Sarcevic, B., Musgrove, E.A., Watts, C.K., Sutherland, R.L., 1997. Estrogen-induced activation of Cdk4 and Cdk2 during G1-S phase progression is accompanied by increased cyclin D1 expression and decreased cyclin-dependent kinase inhibitor association with cyclin E-Cdk2. J. Biol. Chem. 272, 10882–10894.

Prat, A., Parker, J.S., Karginova, O., Fan, C., Livasy, C., Herschkowitz, J.I., He, X., Perou, C.M., 2010. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res. 12, R68.

Ramaswamy, S., Perou, C.M., 2003. DNA microarrays in breast cancer: the promise of personalised medicine. The Lancet 361, 1576–1577.

Rangarajan, A., Hong, S.J., Gifford, A., Weinberg, R.A., 2004. Species- and cell type-specific requirements for cellular transformation. Cancer Cell 6, 171–183.

Reeves, G.K., Travis, R.C., Green, J., Bull, D., Tipper, S., Baker, K., Beral, V., Peto, R., Bell, J., Zelenika, D., Lathrop, M., 2010. Incidence of breast cancer and its subtypes in relation to individual and multiple low-penetrance genetic susceptibility loci. JAMA 304, 426–434.

Reis-Filho, J.S., Tutt, A.N.J., 2008. Triple negative tumours: a critical review. Histopathology 52, 108–118.

Reyal, F., Stransky, N., Bernard-Pierrot, I., Vincent-Salomon, A., de Rycke, Y., Elvin, P., Cassidy, A., Graham, A., Spraggon, C., Désille, Y., Fourquet, A., Nos, C., Pouillart, P., Magdelénat, H., Stoppa-Lyonnet, D., Couturier, J., Sigal-Zafrani, B., Asselain, B., Sastre-Garau, X., Delattre, O., Thiery, J.P., Radvanyi, F., 2005. Visualizing chromosomes as transcriptome correlation maps: evidence of chromosomal domains containing co-expressed genes--a study of 130 invasive ductal breast carcinomas. Cancer Res 65, 1376–1383.

Ricke, W.A., McPherson, S.J., Bianco, J.J., Cunha, G.R., Wang, Y., Risbridger, G.P., 2008. Prostatic hormonal carcinogenesis is mediated by in situ estrogen production and estrogen receptor alpha signaling. FASEB J. 22, 1512–1520.

Rieder, D., Trajanoski, Z., McNally, J.G., 2012. Transcription factories. Front Genet 3, 221.

Riggs, A.D., 1996. Epigenetic mechanisms of gene regulation. Plainview, N.Y: Cold Spring Harbor Laboratory Press.

Rogakou, E.P., Pilch, D.R., Orr, A.H., Ivanova, V.S., Bonner, W.M., 1998. DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. J. Biol. Chem. 273, 5858–5868.

Roh, T., Wei, G., Farrell, C.M., Zhao, K., 2007. Genome-wide prediction of conserved and nonconserved enhancers by histone acetylation patterns. Genome Res. 17, 74–81.

Ross-Innes, C.S., Brown, G.D., Carroll, J.S., 2011. A co-ordinated interaction between CTCF and ER in breast cancer cells. BMC Genomics 12, 593.

Rouleau, M., Patel, A., Hendzel, M.J., Kaufmann, S.H., Poirier, G.G., 2010. PARP inhibition: PARP1 and beyond. Nat. Rev. Cancer 10, 293–301.

Rouquier, S., Taviaux, S., Trask, B.J., Brand-Arpon, V., Engh, G. van den, Demaille, J., Giorgi, D., 1998. Distribution of olfactory receptor genes in the human genome. Nature Genetics 18, 243–250.

Rubio, E.D., Reiss, D.J., Welcsh, P.L., Disteche, C.M., Filippova, G.N., Baliga, N.S., Aebersold, R., Ranish, J.A., Krumm, A., 2008. CTCF physically links cohesin to chromatin. Proc. Natl. Acad. Sci. U.S.A. 105, 8309–8314.

Sachs, R.K., van den Engh, G., Trask, B., Yokota, H., Hearst, J.E., 1995. A random-walk/giant-loop model for interphase chromosomes. Proceedings of the National Academy of Sciences of the United States of America 92, 2710–2714.

Sanders, D.A., Ross-Innes, C.S., Beraldi, D., Carroll, J.S., Balasubramanian, S., 2013. Genome-wide mapping of FOXM1 binding reveals co binding with estrogen receptor alpha in breast cancer cells. Genome Biol. 14, R6.

Santos-Rosa, H., Schneider, R., Bannister, A.J., Sherriff, J., Bernstein, B.E., Emre, N.C.T., Schreiber, S.L., Mellor, J., Kouzarides, T., 2002. Active genes are tri-methylated at K4 of histone H3. Nature 419, 407–411.

Sasieni, P.D., Shelton, J., Ormiston-Smith, N., Thomson, C.S., Silcocks, P.B., 2011. What is the lifetime risk of developing cancer?: the effect of adjusting for multiple primaries. Br. J. Cancer 105, 460–465.

Sawaki, M., Ito, Y., Akiyama, F., Tokudome, N., Horii, R., Mizunuma, N., Takahashi, S., Horikoshi, N., Imai, T., Nakao, A., Kasumi, F., Sakamoto, G., Hatake, K., 2006. High

prevalence of HER-2/neu and p53 overexpression in inflammatory breast cancer. Breast Cancer 13, 172–178.

Schlesinger, Y., Straussman, R., Keshet, I., Farkash, S., Hecht, M., Zimmerman, J., Eden, E., Yakhini, Z., Ben-Shushan, E., Reubinoff, B.E., Bergman, Y., Simon, I., Cedar, H., 2007. Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. Nat. Genet 39, 232–236.

Schmidt, D., Schwalie, P.C., Ross-Innes, C.S., Hurtado, A., Brown, G.D., Carroll, J.S., Flicek, P., Odom, D.T., 2010. A CTCF-independent role for cohesin in tissue-specific transcription. Genome Res 20, 578–588.

Schmidt, M.K., Tollenaar, R.A.E.M., de Kemp, S.R., Broeks, A., Cornelisse, C.J., Smit, V.T.H.B.M., Peterse, J.L., van Leeuwen, F.E., Van't Veer, L.J., 2007. Breast cancer survival and tumor characteristics in premenopausal women carrying the CHEK2*1100delC germline mutation. J. Clin. Oncol. 25, 64–69.

Schmitges, F.W., Prusty, A.B., Faty, M., Stützer, A., Lingaraju, G.M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S., Bunker, R.D., Wirth, U., Bouwmeester, T., Bauer, A., Ly-Hartig, N., Zhao, K., Chan, H., Gu, J., Gut, H., Fischle, W., Müller, J., Thomä, N.H., 2011. Histone methylation by PRC2 is inhibited by active chromatin marks. Mol. Cell 42, 330–341.

Schrader, K.A., Masciari, S., Boyd, N., Wiyrick, S., Kaurah, P., Senz, J., Burke, W., Lynch, H.T., Garber, J.E., Huntsman, D.G., 2008. Hereditary diffuse gastric cancer: association with lobular breast cancer. Fam. Cancer 7, 73–82.

Schrager, C.A., Schneider, D., Gruener, A.C., Tsou, H.C., Peacocke, M., 1998. Clinical and pathological features of breast disease in Cowden's syndrome: an underrecognized syndrome with an increased risk of breast cancer. Hum. Pathol. 29, 47–53.

Shah, S.P., Roth, A., Goya, R., Oloumi, A., Ha, G., et al., 2012. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature 486, 395–399.

Sharp, Z.D., Mancini, M.G., Hinojos, C.A., Dai, F., Berno, V., Szafran, A.T., Smith, K.P., Lele, T.P., Lele, T.T., Ingber, D.E., Mancini, M.A., 2006. Estrogen-receptor-alpha exchange and chromatin dynamics are ligand- and domain-dependent. J. Cell. Sci. 119, 4101–4116.

Shehata, M., Teschendorff, A., Sharp, G., Novcic, N., Russell, I.A., Avril, S., Prater, M., Eirew, P., Caldas, C., Watson, C.J., 2012. Phenotypic and functional characterisation of the luminal cell hierarchy of the mammary gland. Breast Cancer Research 14, R134.

Shi, Yujiang, Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., Shi, Yang, 2004. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell 119, 941–953.

Shilatifard, A., 2012. The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. Annu. Rev. Biochem. 81, 65–95.

Shopland, L.S., Lynch, C.R., Peterson, K.A., Thornton, K., Kepper, N., Hase, J. von, Stein, S., Vincent, S., Molloy, K.R., Kreth, G., Cremer, C., Bult, C.J., O'Brien, T.P., 2006. Folding and organization of a contiguous chromosome region according to the gene distribution pattern in primary genomic sequence. J Cell Biol 174, 27–38.

Simic, R., Lindstrom, D.L., Tran, H.G., Roinick, K.L., Costa, P.J., Johnson, A.D., Hartzog, G.A., Arndt, K.M., 2003. Chromatin remodeling protein Chd1 interacts with transcription elongation factors and localizes to transcribed genes. EMBO J. 22, 1846–1856.

Simon, J.A., Kingston, R.E., 2009. Mechanisms of Polycomb gene silencing: knowns and unknowns. Nature Reviews Molecular Cell Biology 10, 697–708.

Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B., de Laat, W., 2006. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat. Genet. 38, 1348–1354.

Smalley, M., Ashworth, A., 2003. Stem cells and breast cancer: A field in transit. Nat. Rev. Cancer 3, 832–844.

Solovei, I., Kreysing, M., Lanctôt, C., Kösem, S., Peichl, L., Cremer, T., Guck, J., Joffe, B., 2009. Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution. Cell 137, 356–368.

Sørlie, T., 2004. Molecular portraits of breast cancer: tumour subtypes as distinct disease entities. European Journal of Cancer 40, 2667–2675.

Sørlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Thorsen, T., Quist, H., Matese, J.C., Brown, P.O., Botstein, D., Lønning, P.E., Børresen-Dale, A.L., 2001. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc. Natl. Acad. Sci. U.S.A. 98, 10869–10874.

Sorlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J.S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S., Demeter, J., Perou, C.M., Lønning, P.E., Brown, P.O., Børresen-Dale, A.-L., Botstein, D., 2003. Repeated observation of breast tumor

subtypes in independent gene expression data sets. Proc. Natl. Acad. Sci. U.S.A. 100, 8418–8423.

Sørlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J.S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S., Demeter, J., Perou, C.M., Lønning, P.E., Brown, P.O., Børresen-Dale, A.-L., Botstein, D., 2003. Repeated observation of breast tumor subtypes in independent gene expression data sets. PNAS 100, 8418–8423.

Sotiriou, C., Pusztai, L., 2009. Gene-expression signatures in breast cancer. N. Engl. J. Med 360, 790–800.

Spike, B.T., Engle, D.D., Lin, J.C., Cheung, S.K., La, J., Wahl, G.M., 2012. A mammary stem cell population identified and characterized in late embryogenesis reveals similarities to human breast cancer. Cell Stem Cell 10, 183–197.

Spilianakis, C.G., Lalioti, M.D., Town, T., Lee, G.R., Flavell, R.A., 2005. Interchromosomal associations between alternatively expressed loci. Nature 435, 637–645.

Splinter, E., Heath, H., Kooren, J., Palstra, R.-J., Klous, P., Grosveld, F., Galjart, N., de Laat, W., 2006. CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. Genes Dev 20, 2349–2354.

Splinter, E., Wit, E. de, Nora, E.P., Klous, P., Werken, H.J.G. van de, Zhu, Y., Kaaij, L.J.T., IJcken, W. van, Gribnau, J., Heard, E., Laat, W. de, 2011. The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA. Genes Dev. 25, 1371–1383.

Sproul, D., Gilbert, N., Bickmore, W.A., 2005. The role of chromatin structure in regulating the expression of clustered genes. Nat. Rev. Genet 6, 775–781.

Sproul, D., Kitchen, R.R., Nestor, C.E., Dixon, J.M., Sims, A.H., Harrison, D.J., Ramsahoye, B.H., Meehan, R.R., 2012. Tissue of origin determines cancer-associated CpG island promoter hypermethylation patterns. Genome Biol. 13, R84.

Sproul, D., Nestor, C., Culley, J., Dickson, J.H., Dixon, J.M., Harrison, D.J., Meehan, R.R., Sims, A.H., Ramsahoye, B.H., 2011. Transcriptionally repressed genes become aberrantly methylated and distinguish tumors of different lineages in breast cancer. Proceedings of the National Academy of Sciences 108, 4364 –4369.

Squazzo, S.L., Costa, P.J., Lindstrom, D.L., Kumer, K.E., Simic, R., Jennings, J.L., Link, A.J., Arndt, K.M., Hartzog, G.A., 2002. The Paf1 complex physically and functionally associates with transcription elongation factors in vivo. EMBO J. 21, 1764–1774.

Stahlberg, C., Pedersen, A.T., Lynge, E., Andersen, Z.J., Keiding, N., Hundrup, Y.A., Obel, E.B., Ottesen, B., 2004. Increased risk of breast cancer following different

regimens of hormone replacement therapy frequently used in Europe. Int. J. Cancer 109, 721–727.

Statham, A.L., Robinson, M.D., Song, J.Z., Coolen, M.W., Stirzaker, C., Clark, S.J., 2012. Bisulfite sequencing of chromatin immunoprecipitated DNA (BisChIP-seq) directly informs methylation status of histone-modified DNA. Genome Res. 22, 1120–1127.

Stedman, W., Kang, H., Lin, S., Kissil, J.L., Bartolomei, M.S., Lieberman, P.M., 2008. Cohesins localize with CTCF at the KSHV latency control region and at cellular c-myc and H19/Igf2 insulators. EMBO J. 27, 654–666.

Steensel, B. van, Henikoff, S., 2000. Identification of in vivo DNA targets of chromatin proteins using tethered Dam methyltransferase. Nature Biotechnology 18, 424–428.

Stephens, P.J., Tarpey, P.S., Davies, H., Van Loo, P., Greenman, C., et al., 2012. The landscape of cancer genes and mutational processes in breast cancer. Nature 486, 400–404.

Stransky, N., Vallot, C., Reyal, F., Bernard-Pierrot, I., de Medina, S.G.D., Segraves, R., de Rycke, Y., Elvin, P., Cassidy, A., Spraggon, C., Graham, A., Southgate, J., Asselain, B., Allory, Y., Abbou, C.C., Albertson, D.G., Thiery, J.P., Chopin, D.K., Pinkel, D., Radvanyi, F., 2006. Regional copy number-independent deregulation of transcription in cancer. Nat Genet 38, 1386–1396.

Sutherland, H., Bickmore, W.A., 2009. Transcription factories: gene expression in unions? Nat. Rev. Genet 10, 457–466.

Tan, M., Yao, J., Yu, D., 1997. Overexpression of the c-erbB-2 gene enhanced intrinsic metastasis potential in human breast cancer cells without increasing their transformation abilities. Cancer Res. 57, 1199–1205.

Tan, S.-H., Wolff, A.C., 2007. Luteinizing hormone-releasing hormone agonists in premenopausal hormone receptor-positive breast cancer. Clin. Breast Cancer 7, 455–464.

Teschendorff, A.E., Miremadi, A., Pinder, S.E., Ellis, I.O., Caldas, C., 2007. An immune response gene expression module identifies a good prognosis subtype in estrogen receptor negative breast cancer. Genome Biol. 8, R157.

Teyssier, C., Le Romancer, M., Sentis, S., Jalaguier, S., Corbo, L., Cavaillès, V., 2010. Protein arginine methylation in estrogen signaling and estrogen-related cancers. Trends Endocrinol. Metab. 21, 181–189.

Thompson, D., Duedal, S., Kirner, J., McGuffog, L., Last, J., Reiman, A., Byrd, P., Taylor, M., Easton, D.F., 2005. Cancer Risks and Mortality in Heterozygous ATM Mutation Carriers. JNCI J Natl Cancer Inst 97, 813–822.

Thorstenson, Y.R., Roxas, A., Kroiss, R., Jenkins, M.A., Yu, K.M., Bachrich, T., Muhr, D., Wayne, T.L., Chu, G., Davis, R.W., Wagner, T.M.U., Oefner, P.J., 2003. Contributions of ATM Mutations to Familial Breast and Ovarian Cancer. Cancer Res 63, 3325–3333.

Tiwari, V.K., Cope, L., McGarvey, K.M., Ohm, J.E., Baylin, S.B., 2008. A novel 6C assay uncovers Polycomb-mediated higher order chromatin conformations. Genome Research 18, 1171–1179.

Tolhuis, B., Blom, M., Kerkhoven, R.M., Pagie, L., Teunissen, H., Nieuwland, M., Simonis, M., de Laat, W., van Lohuizen, M., van Steensel, B., 2011. Interactions among Polycomb domains are guided by chromosome architecture. PLoS Genet. 7, e1001343.

Turner, N., Tutt, A., Ashworth, A., 2005. Targeting the DNA repair defect of BRCA tumours. Current Opinion in Pharmacology 5, 388–393.

Tynan, S.H., Lundeen, S.G., Allan, G.F., 2004. Cell type-specific bidirectional regulation of the glucocorticoid-induced leucine zipper (GILZ) gene by estrogen. The Journal of Steroid Biochemistry and Molecular Biology 91, 225–239.

Vallot, C., Stransky, N., Bernard-Pierrot, I., Herault, A., Zucman-Rossi, J., Chapeaublanc, E., Vordos, D., Laplanche, A., Benhamou, S., Lebret, T., Southgate, J., Allory, Y., Radvanyi, F., 2010. A Novel Epigenetic Phenotype Associated With the Most Aggressive Pathway of Bladder Tumor Progression. JNCI Journal of the National Cancer Institute 103, 47–60.

Van de Wiel, M.A., Kim, K.I., Vosse, S.J., van Wieringen, W.N., Wilting, S.M., Ylstra, B., 2007. CGHcall: calling aberrations for array CGH tumor profiles. Bioinformatics 23, 892 –894.

Van den Engh, G., Sachs, R., Trask, B.J., 1992. Estimating genomic distance from DNA sequence location in cell nuclei by a random walk model. Science 257, 1410–1412.

Vargo-Gogola, T., Rosen, J.M., 2007. Modelling breast cancer: one size does not fit all. Nat. Rev. Cancer 7, 659–672.

Vernimmen, D., Gobbi, M.D., Sloane-Stanley, J.A., Wood, W.G., Higgs, D.R., 2007. Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. EMBO J 26, 2041–2051.

Versteeg, R., van Schaik, B.D.C., van Batenburg, M.F., Roos, M., Monajemi, R., Caron, H., Bussemaker, H.J., van Kampen, A.H.C., 2003. The human transcriptome map reveals extremes in gene density, intron length, GC content, and repeat pattern for domains of highly and weakly expressed genes. Genome Res 13, 1998–2004.

Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., Afzal, V., Ren, B., Rubin, E.M., Pennacchio, L.A., 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature 457, 854–858.

Visvanathan, K., Chlebowski, R.T., Hurley, P., Col, N.F., Ropka, M., Collyar, D., Morrow, M., Runowicz, C., Pritchard, K.I., Hagerty, K., Arun, B., Garber, J., Vogel, V.G., Wade, J.L., Brown, P., Cuzick, J., Kramer, B.S., Lippman, S.M., 2009. American society of clinical oncology clinical practice guideline update on the use of pharmacologic interventions including tamoxifen, raloxifene, and aromatase inhibition for breast cancer risk reduction. J. Clin. Oncol. 27, 3235–3258.

Vogel, M.J., Guelen, L., de Wit, E., Hupkes, D.P., Lodén, M., Talhout, W., Feenstra, M., Abbas, B., Classen, A.-K., van Steensel, B., 2006. Human heterochromatin proteins form large domains containing KRAB-ZNF genes. Genome Research 16, 1493 –1504.

Volpi, E.V., Chevret, E., Jones, T., Vatcheva, R., Williamson, J., Beck, S., Campbell, R.D., Goldsworthy, M., Powis, S.H., Ragoussis, J., Trowsdale, J., Sheer, D., 2000. Large-scale chromatin organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei. J. Cell. Sci. 113 (Pt 9), 1565–1576.

Vousden, K.H., Prives, C., 2005. P53 and prognosis: new insights and further complexity. Cell 120, 7–10.

Waddington, C.H., n.d. The Strategy of the Genes. Allen & Unwin, London 1957.

Wang, H., Huang, Z.Q., Xia, L., Feng, Q., Erdjument-Bromage, H., Strahl, B.D., Briggs, S.D., Allis, C.D., Wong, J., Tempst, P., Zhang, Y., 2001. Methylation of histone H4 at arginine 3 facilitating transcriptional activation by nuclear hormone receptor. Science 293, 853–857.

Wang, K.C., Yang, Y.W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., Lajoie, B.R., Protacio, A., Flynn, R.A., Gupta, R.A., Wysocka, J., Lei, M., Dekker, J., Helms, J.A., Chang, H.Y., 2011. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature 472, 120–124.

Wang, Y.-Q., Yan, Q., Zhang, J.-R., Li, S.-D., Yang, Y.-X., Wan, X.-P., 2012. Epigenetic inactivation of BRCA1 through promoter hypermethylation in ovarian cancer progression. J. Obstet. Gynaecol. Res.

Watson, C.J., Khaled, W.T., 2008. Mammary development in the embryo and adult: a journey of morphogenesis and commitment. Development 135, 995–1003.

Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Pääbo, S., Rebhan, M., Schübeler, D., 2007. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat. Genet. 39, 457–466.

Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., Yahata, K., Imamoto, F., Aburatani, H., Nakao, M., Imamoto, N., Maeshima, K., Shirahige, K., Peters, J.-M., 2008. Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature 451, 796–801.

Whirl-Carrillo, M., McDonagh, E.M., Hebert, J.M., Gong, L., Sangkuhl, K., Thorn, C.F., Altman, R.B., Klein, T.E., 2012. Pharmacogenomics knowledge for personalized medicine. Clin. Pharmacol. Ther. 92, 414–417.

Widschwendter, M., Fiegl, H., Egle, D., Mueller-Holzner, E., Spizzo, G., Marth, C., Weisenberger, D.J., Campan, M., Young, J., Jacobs, I., Laird, P.W., 2007. Epigenetic stem cell signature in cancer. Nat. Genet 39, 157–158.

Williamson, I., Eskeland, R., Lettice, L.A., Hill, A.E., Boyle, S., Grimes, G.R., Hill, R.E., Bickmore, W.A., 2012. Anterior-posterior differences in HoxD chromatin topology in limb development. Development 139, 3157–3167.

Winkler, D.D., Luger, K., 2011. The histone chaperone FACT: structural insights and mechanisms for nucleosome reorganization. J. Biol. Chem. 286, 18369–18374.

Wolff, A.C., Hammond, M.E.H., Schwartz, J.N., Hagerty, K.L., Allred, D.C., Cote, R.J., Dowsett, M., Fitzgibbons, P.L., Hanna, W.M., Langer, A., McShane, L.M., Paik, S., Pegram, M.D., Perez, E.A., Press, M.F., Rhodes, A., Sturgeon, C., Taube, S.E., Tubbs, R., Vance, G.H., van de Vijver, M., Wheeler, T.M., Hayes, D.F., 2007. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. Arch. Pathol. Lab. Med. 131, 18–43.

Wright, D.E., Wang, C.-Y., Kao, C.-F., 2012. Histone ubiquitylation and chromatin dynamics. Front. Biosci. 17, 1051–1078.

Wu, H., D'Alessio, A.C., Ito, S., Xia, K., Wang, Z., Cui, K., Zhao, K., Sun, Y.E., Zhang, Y., 2011. Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. Nature 473, 389–393.

Würtele, H., Chartrand, P., 2006. Genome-wide scanning of HoxB1-associated loci in mouse ES cells using an open-ended Chromosome Conformation Capture methodology. Chromosome Res. 14, 477–495.

Wysocka, J., Allis, C.D., Coonrod, S., 2006a. Histone arginine methylation and its dynamic regulation. Front. Biosci. 11, 344–355.

Wysocka, J., Swigut, T., Xiao, H., Milne, T.A., Kwon, S.Y., Landry, J., Kauer, M., Tackett, A.J., Chait, B.T., Badenhorst, P., Wu, C., Allis, C.D., 2006b. A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. Nature 442, 86–90.

Xu, W., Chen, H., Du, K., Asahara, H., Tini, M., Emerson, B.M., Montminy, M., Evans, R.M., 2001. A transcriptional switch mediated by cofactor methylation. Science 294, 2507–2511.

Xu, W., Cho, H., Kadam, S., Banayo, E.M., Anderson, S., Yates, J.R., 3rd, Emerson, B.M., Evans, R.M., 2004. A methylation-mediator complex in hormone signaling. Genes Dev. 18, 144–156.

Yokota, H., van den Engh, G., Hearst, J.E., Sachs, R.K., Trask, B.J., 1995. Evidence for the organization of chromatin in megabase pair-sized loops arranged along a random walk path in the human G0/G1 interphase nucleus. J. Cell Biol 130, 1239–1249.

Yokoyama, A., Wang, Z., Wysocka, J., Sanyal, M., Aufiero, D.J., Kitabayashi, I., Herr, W., Cleary, M.L., 2004. Leukemia proto-oncoprotein MLL forms a SET1-like histone methyltransferase complex with menin to regulate Hox gene expression. Mol. Cell. Biol. 24, 5639–5649.

Yuan, G.-C., Liu, Y.-J., Dion, M.F., Slack, M.D., Wu, L.F., Altschuler, S.J., Rando, O.J., 2005. Genome-scale identification of nucleosome positions in S. cerevisiae. Science 309, 626–630.

Zhang, X., Bolt, M., Guertin, M.J., Chen, W., Zhang, S., Cherrington, B.D., Slade, D.J., Dreyton, C.J., Subramanian, V., Bicker, K.L., Thompson, P.R., Mancini, M.A., Lis, J.T., Coonrod, S.A., 2012. Peptidylarginine deiminase 2-catalyzed histone H3 arginine 26 citrullination facilitates estrogen receptor α target gene activation. Proc. Natl. Acad. Sci. U.S.A. 109, 13331–13336.

Zhang, Y., Liang, J., Li, Y., Xuan, C., Wang, F., Wang, D., Shi, L., Zhang, D., Shang, Y., 2010. CCCTC-binding factor acts upstream of FOXA1 and demarcates the genomic response to estrogen. J. Biol. Chem. 285, 28604–28613.

Zhang, Y., Reinberg, D., 2001. Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. Genes Dev. 15, 2343–2360.

Zhu, X., Leav, I., Leung, Y.-K., Wu, M., Liu, Q., Gao, Y., McNeal, J.E., Ho, S.-M., 2004. Dynamic regulation of estrogen receptor-beta expression by DNA methylation during prostate cancer development and metastasis. Am. J. Pathol. 164, 2003–2012.

Zwart, W., Theodorou, V., Kok, M., Canisius, S., Linn, S., Carroll, J.S., 2011. Oestrogen receptor-co-factor-chromatin specificity in the transcriptional regulation of breast cancer. EMBO J. 30, 4764–4776.

Chapter 9

Appendices

Appendices

Legends for appendices on data CD

Appendix I

DNA Copy Number Profiles for 48 Breast Cancer Cell Lines

The log2 (sample:ref) ratios of the clones were plotted against chromosomal position. Regions of loss (green), gain (red), and no change (black) as determined by CGHcall (van de Wiel et al., 2007).

Appendix II

DNA Copy Number Profiles in 356 Breast Tumours

The log2 (sample:ref) ratios of the clones were plotted against chromosomal position. Regions of loss (green), gain (red), and no change (black) as determined by CGHcall (van de Wiel et al., 2007).

Appendix III

Breast tumour TCS Maps for all chromosomes

Transcriptome correlation maps generated by plotting all TCS values against chromosome position in megabases (Mb) using data for 356 tumours (Jönsson et al., 2010). The dotted line represents the significance threshold. Genes above this line are considered to have expression profiles that are considerably correlated with their neighbours.

Appendix IV

Analysis of gene expression changes in tumours relative to normal tissue for 45 RER domains derived from breast tumour analysis.

Mean centred z scores were calculated for all genes in the region. The median z score for each sample was used to plot the distribution of these scores as boxplots and to compare

scores between samples. Significant changes in expression were determined by the Wilcox test ($p<0.05$).

Appendix V

Breast cancer cell line TCS Maps for all chromosomes

Transcriptome correlation maps generated by plotting all TCS values against chromosome position in megabases (Mb) using data for 48 cell lines (Neve et al., 2006). The dotted line represents the significance threshold. Genes above this line are considered to have expression profiles that are considerably correlated with their neighbours.

Appendix VI

26 overlapping RER domains between breast tumour and cell lines

The BioMart function at Ensembl.org was used to extract the Ensembl Ids and Associated Gene Names for all genes within the coordinates for the 26 overlapping RER domains.

Appendix VII

Analysis of gene expression changes in tumours relative to normal tissue for 26 overlapping RER domains between tumour and cell line analysis.

Mean centred z scores were calculated for all genes in the region. The median z score for each sample was used to plot the distribution of these scores as boxplots and to compare scores between samples. Significant changes in expression were determined by the Wilcox test ($p<0.05$).

Appendix VIII

Unsupervised cluster analysis of genes in breast cancer RER domains

Heat maps of the 45 tumour data derived RER domains. Each tumour is identified by a color-coded matrix below the dendrogram: basal (red), luminal (blue), ERBB2 (purple), normal-like (green). Red/green indicate an increase/decrease in gene expression relative to the universal mean for each gene. The values were mean centred and the colours scaled from the max (+) & min (-) standard deviations for each heat map.

Heat maps of the 71 cell line data derived RER domains. Each cell line is identified by a color-coded matrix below the dendrogram: basal (red), and luminal (blue). Red/green indicate an increase/decrease in gene expression relative to the universal mean for each gene. The values were mean centred and the colours scaled from the max (+) & min (-) standard deviations for each heat map.

Appendix IX

GOrilla Gene Ontology analysis of breast cancer RER genes

Top GO terms for 26 overlapping RER domains in breast cancer

Appendix X

Comparing the localisations of breast cancer RER domains with RIDGEs/antiRIDGEs

Table 1: The number of RER domains overlapping with Regions of Increased Gene Expresion (RIDGEs) in randomly generated regions (expected) and in RER domains (observed). Fisher's exact test (p=0.395) showed there was no significant correlation between RER domains and RIDGE's above what would be expected by chance.

Table 2: The number of RER domains overlapping with anti-RIDGEs in randomly generated regions (expected) and in RER domains (observed). Fisher's exact test (p=0.1) showed there was no significant correlation between RER domains and anti-RIDGE's above what would be expected by chance.

Appendix XI

Identifying copy number independent regions of transcriptome bias in liver heaptoblastoma

Copy number profiles for 25 hepatoblastoma samples: The log2 (sample:ref) ratios of the clones were plotted against chromosomal position. Regions of loss (green), gain (red), and no change (black) as determined by CGHcall (van de Wiel et al., 2007).

Heptoblastoma TCS Maps for all chromosomes: Transcriptome correlation maps generated by plotting all TCS values against chromosome position in megabases (Mb) using data for 25tumours. The dotted line represents the significance threshold. Genes

above this line are considered to have expression profiles that are considerably correlated with their neighbours.